

STIC-Biotech/ChemLib

174942

From: Ramirez, Delia
Sent: Thursday, December 22, 2005 3:53 PM
To: STIC-Biotech/ChemLib
Subject: 10/602220

Hi,

I would like to request the following searches:

1. SEQ ID NO: 15, 16 and 29 in the nucleic acid databases (commercial and interference)
2. an alignment of SEQ ID NO:16 and 29
3. SEQ ID NO:16 and 29 in the protein databases (commercial & interference)

Thank you very much,

Delia M. Ramirez, Ph.D.
Patent Examiner
Recombinant Enzymes-Art Unit 1652
USPTO
400 Dulany Street, Remsen Bldg., 2D74, Mail room 2C70
Alexandria, VA 22314
(571) 272-0938
delia.ramirez@uspto.gov

DEC 23 2005
USPTO

Searcher: _____
Searcher Phone: _____
Date Searcher Picked up: _____
Date completed: 4/3/06
Searcher Prep Time: _____
Online Time: _____

Type of Search
NA# _____ AA# _____
S/L: _____ Oligomer: _____
Encode/Transl: _____
Structure #: _____ Text: _____
Inventor: _____ Litigation: _____

Vendors and cost where applicable
STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other (Specify): _____

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 31, 2005, 15:41:01 ; Search time 6619 Seconds
(without alignments)
10769.253 Million cell updates/sec

Title: US-10-602-220-15
Perfect score: 1254
Sequence: 1 atgcagctgaggaccaga.....ctgtttgttcagctataa 1254

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 588141 seqs, 28421725653 residues
Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:

- 1: gb_ba.*
- 2: gb_in.*
- 3: gb_env.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pr.*
- 9: gb_ro.*
- 10: gb_sts.*
- 11: gb_sy.*
- 12: gb_un.*
- 13: gb_vi.*
- 14: gb_htg.*
- 15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1254	100.0	1254	6	AR653698 Sequence
2	1254	100.0	1254	6	AR654966 Sequence
3	1252.4	99.9	1266	6	AR653696 Sequence
4	1252.4	99.9	1266	6	AR654964 Sequence
5	1252.4	99.9	1278	6	AR653694 Sequence
6	1252.4	99.9	1278	6	AR654962 Sequence
7	1252.4	99.9	1284	6	AR653697 Sequence
8	1252.4	99.9	1284	6	AR654965 Sequence
9	1252.4	99.9	1290	6	CS135670 Sequence
10	1252.4	99.9	1290	6	AR653692 Sequence
11	1252.4	99.9	1290	6	AR654960 Sequence
12	1252.4	99.9	1290	8	BT006864 Homo sapi
13	1252.4	99.9	1290	11	AY889631 Synthetic
14	1252.4	99.9	1290	11	AY889632 Synthetic
15	1252.4	99.9	1290	11	AY892090 Synthetic
16	1252.4	99.9	1290	11	BT007835 Synthetic
17	1252.4	99.9	1296	6	AR653695 Sequence
18	1252.4	99.9	1296	6	AR654963 Sequence

19	1252.4	99.9	1306	6	BD182986	BD182986 Pharmaceu
20	1252.4	99.9	1306	6	BD185988	BD185988 Pharmaceu
21	1252.4	99.9	1308	6	AR653693	AR653693 Sequence
22	1252.4	99.9	1308	6	AR654961	AR654961 Sequence
23	1252.4	99.9	1312	8	BC002689	BC002689 Homo sapi
24	1252.4	99.9	1319	6	AX840234	AX840234 Sequence
25	1252.4	99.9	1319	8	HSAGALAR	XO5790 Human mRNA
26	1252.4	99.9	1343	6	AR101984	AR101984 Sequence
27	1252.4	99.9	1343	6	BD064229	BD064229 Therapy f
28	1252.4	99.9	1343	6	BD247365	BD247365 Treatment
29	1252.4	99.9	1343	6	CS105027	CS105027 Sequence
30	1252.4	99.9	1343	6	AR234439	AR234439 Sequence
31	1252.4	99.9	1343	6	AR322462	AR322462 Sequence
32	1252.4	99.9	1343	6	AR372138	AR372138 Sequence
33	1252.4	99.9	1350	6	CQ899396	CQ899396 Sequence
34	1252.4	99.9	1350	6	CS033394	CS033394 Sequence
35	1252.4	99.9	1350	6	CS035758	CS035758 Sequence
36	1252.4	99.9	1350	6	CS042346	CS042346 Sequence
37	1252.4	99.9	1350	6	CS044710	CS044710 Sequence
38	1252.4	99.9	1390	6	AX780273	AX780273 Sequence
39	1252.4	99.9	1393	6	I30313	I30313 Sequence 6
40	1252.4	99.9	1393	6	AX030659	AX030659 Sequence
41	1252.4	99.9	1447	6	BD134457	BD134457 Human nuc
42	1252.4	99.9	1447	6	AX017286	AX017286 Sequence
43	1252.4	99.9	1447	6	AX524990	AX524990 Sequence
44	1252	99.8	1272	6	AR653699	AR653699 Sequence
45	1252	99.8	1272	6	AR654967	AR654967 Sequence

ALIGNMENTS

RESULT 1
AR653698
LOCUS AR653698 1254 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 15 from patent US 6887696.
ACCESSION AR653698
VERSION AR653698.1 GI:67584283
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1254)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITILE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 15 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source
1..1254
/organism="unknown"
/mol_type="genomic DNA"

Query Match	Best Local Similarity	Matches 1254;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGCCTTCTCTGGCC	60							
Db	1	ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGCCTTCTCTGGCC	60							
QY	61	CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACATGATGGCAAGGACGCTT	120							
Db	61	CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACATGATGGCAAGGACGCTT	120							
QY	121	ACCATGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACCTCCAGGAGGCCA	180							
Db	121	ACCATGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACCTCCAGGAGGCCA	180							
QY	181	GATTCTCTCATCAGTGAGAGGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC	240							
Db	181	GATTCTCTCATCAGTGAGAGGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC	240							


```
QY 781 GGTGGATGACCCAGATATGTTAGTGATTTGGGCACTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGATGACCCAGATATGTTAGTGATTTGGCACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTTATGCTATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTTATGCTATGACCTC 900
QY 901 CGACATCATAGCCCTCAAGCAAGCTCTCTTTTCCAGGATAAGGACGTAATTTGCCATCAAT 960
Db 901 CGACATCATAGCCCTCAAGCAAGCTCTCTTTTCCAGGATAAGGACGTAATTTGCCATCAAT 960
QY 961 CAGACCCCTTGGCAAGCAAGGTAACAGCTTACAGCTTTAGACAGGAGACAACTTTGAAGTGTG 1020
Db 961 CAGACCCCTTGGCAAGCAAGGTAACAGCTTACAGCTTTAGACAGGAGACAACTTTGAAGTGTG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATATAACCGCAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATATAACCGCAGGAGATTGGT 1080
QY 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTCTGAATCCCT 1140
Db 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTCTGAATCCCT 1140
QY 1141 GCCTGCTTATCATCAGACCTCTCCCTGTGAAAGGAAGCTAGGGTCTTATGAATGGACT 1200
Db 1141 GCCTGCTTATCATCAGACCTCTCCCTGTGAAAGGAAGCTAGGGTCTTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTATATAA 1254
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTATATAA 1254
```

```
RESULT 3
AR653696
LOCUS AR653696 1266 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 11 from patent US 6887696.
ACCESSION AR653696
VERSION AR653696.1 GI:67584281
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 1266)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 11 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source
1..1266
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 99.9%; Score 1252.4; DB 6; Length 1266;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCGCTTCCGCTGGCC 60
Db 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCGCTTCCGCTGGCC 60
QY 61 CTGCTTTCTTGGACATCCCTGGGCTTAGAGCACTGGACAAATGGATTGGCAAGGAGCGCT 120
Db 61 CTGCTTTCTTGGACATCCCTGGGCTTAGAGCACTGGACAAATGGATTGGCAAGGAGCGCT 120
QY 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCAATGTGCAACCTTGAATGCCAGGAGAGCCA 180
Db 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCAATGTGCAACCTTGAATGCCAGGAGAGCCA 180
QY 181 GATTCTTCATCAGTGAAGAGCTCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAAGGC 240
Db 181 GATTCTTCATCAGTGAAGAGCTCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAAGGC 240
```

```
QY 241 TGAAGGATGACGGTTATGAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCCAAAGA 300
Db 241 TGAAGGATGACGGTTATGAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCCAAAGA 300
QY 301 GATTTCAGAGGAGAGCTTTCAGGAGAGCCCTCAGCGCTTCTCATGGGATTCGCGAGCTA 360
Db 301 GATTTCAGAGGAGAGCTTTCAGGAGAGCCCTCAGCGCTTCTCATGGGATTCGCGAGCTA 360
QY 361 GCTAAATTATGTTTCAACAGCAAGGACTGAAGCTAGGGATTTATGAGATGTTGGAATAAAA 420
Db 361 GCTAAATTATGTTTCAACAGCAAGGACTGAAGCTAGGGATTTATGAGATGTTGGAATAAAA 420
QY 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATATACTACGACATTGATGCCAGACCTTTGCT 480
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATATACTACGACATTGATGCCAGACCTTTGCT 480
QY 481 GACTGGGGAGTAGATCTGCTAAATTTGATGTTTGTCTGTGACAGATTTGGAATAATTTG 540
Db 481 GACTGGGGAGTAGATCTGCTAAATTTGATGTTTGTCTGTGACAGATTTGGAATAATTTG 540
QY 541 GCAGATGGTTTATAAGCACATGCTTGGCCCTGAATAGGACTGGCAGAAAGCTTGTGTAC 600
Db 541 GCAGATGGTTTATAAGCACATGCTTGGCCCTGAATAGGACTGGCAGAAAGCTTGTGTAC 600
QY 601 TCCGTGTAGTGGCTCTTTTATATGTCGCCCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCGTGTAGTGGCTCTTTTATATGTCGCCCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCAGCTGGCGAAATTTTGTGTCATTTGATGATTCCTGGAAGATTAAG 720
Db 661 CAGTACTGCAATCAGCTGGCGAAATTTTGTGTCATTTGATGATTCCTGGAAGATTAAG 720
QY 721 AGTATCTTGACCTGGACATCTTTTAAACAGGAGAGAAATTTGTCATGTTGTCACCAAGG 780
Db 721 AGTATCTTGACCTGGACATCTTTTAAACAGGAGAGAAATTTGTCATGTTGTCACCAAGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTATGTCGCAACTTTGGCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAATGACCCAGATATGTTAGTATGTCGCAACTTTGGCTCAGCTGGAATCAGCAA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
QY 901 CGACATCATAGCCCTCAAGCAAGCTCTCTTCCAGGATAAGGACGTAATTTGCCATCAAT 960
Db 901 CGACATCATAGCCCTCAAGCAAGCTCTCTTCCAGGATAAGGACGTAATTTGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTAACAGCTTACAGCTTTAGACAGGAGACAACTTTGAAGTGTG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTAACAGCTTACAGCTTTAGACAGGAGACAACTTTGAAGTGTG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATATAACCGCAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATATAACCGCAGGAGATTGGT 1080
QY 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTCTGAATCCCT 1140
Db 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTCTGAATCCCT 1140
QY 1141 GCCTGCTTATCATCAGACCTCTCCCTGTGAAAGGAAGCTAGGGTCTTATGAATGGACT 1200
Db 1141 GCCTGCTTATCATCAGACCTCTCCCTGTGAAAGGAAGCTAGGGTCTTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTATATAA 1254
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTATATAA 1254
RESULT 4
AR654964
LOCUS AR654964 1266 bp DNA linear PAT 13-JUN-2005
```

```
DEFINITION Sequence 11 from patent US 6890748.
ACCESSION AR654964
VERSION AR654964.1 GI:67586503
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1266)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 11 10-MAY-2005;
LARGE SCALE Biology Corporation; Vacaville, CA
FEATURES
source
1..1266
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 99.9%; Score 1252.4; DB 6; Length 1266;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCGCTTCCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCGCTTCCTGGCC 60
QY 61 CTGCTTTCCTGGGACATCCCTGGGCTAGAGCACTGGACAAATGGATGGCAAGGACGCT 120
DB 61 CTGCTTTCCTGGGACATCCCTGGGCTAGAGCACTGGACAAATGGATGGCAAGGACGCT 120
QY 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
DB 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
QY 181 GATTCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAGGC 240
DB 181 GATTCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAGGC 240
QY 241 TGGAGGATGACGGTTATGATGATACCTCTGCAATGATGATGATGATGATGATGATGAT 300
DB 241 TGGAGGATGACGGTTATGATGATACCTCTGCAATGATGATGATGATGATGATGATGAT 300
QY 301 GATTGAGAGGACAGCTTCAGGAGACCCCTGAGGCTTCCTCATGGATTCGCCAGCTA 360
DB 301 GATTGAGAGGACAGCTTCAGGAGACCCCTGAGGCTTCCTCATGGATTCGCCAGCTA 360
QY 361 GCTAATTTATGTTTACAGCAAGGACTGAAGCTAGGAGATTTATGACATGTTGGAATATA 420
DB 361 GCTAATTTATGTTTACAGCAAGGACTGAAGCTAGGAGATTTATGACATGTTGGAATATA 420
QY 421 ACCTGGCAGGCTTCCTGGGATTTGGATATCTACGACATGATGATGATGATGATGATGAT 480
DB 421 ACCTGGCAGGCTTCCTGGGATTTGGATATCTACGACATGATGATGATGATGATGATGAT 480
QY 481 GACTGGGAGTACATCTGCTAAATTTGATGTTGTTTACGTGACAGTTTGGAAATTTG 540
DB 481 GACTGGGAGTACATCTGCTAAATTTGATGTTGTTTACGTGACAGTTTGGAAATTTG 540
QY 541 GCAGATGTTTAAAGACATGTCCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
DB 541 GCAGATGTTTAAAGACATGTCCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTCATGATGATGATGATGATGATGATGAT 720
DB 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATTCATGATGATGATGATGATGATGATGAT 720
QY 721 AGTATCTTGACATGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACACAGGG 780
DB 721 AGTATCTTGACATGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACACAGGG 780
```

```
QY 781 GGTTCGAATGACCCAGATATGTTAGTGATTGGCAACTTTGGCTCAGCTGGAAATCAGCAA 840
DB 781 GGTTCGAATGACCCAGATATGTTAGTGATTGGCAACTTTGGCTCAGCTGGAAATCAGCAA 840
QY 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
DB 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
QY 901 CGACATCAGCCCTCAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCGCATCAAT 960
DB 901 CGACATCAGCCCTCAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCGCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTAGCTATGATATAAACCAGGAGAGATTGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTAGCTATGATATAAACCAGGAGAGATTGGT 1080
QY 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCTGGGTAAAGAGTGGCCTGTAATCCT 1140
DB 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCTGGGTAAAGAGTGGCCTGTAATCCT 1140
QY 1141 GCCTGCTTCATCACACAGCTTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
DB 1141 GCCTGCTTCATCACACAGCTTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 1201 TCAAGTTAAGAGTCACATAATCCACAGGACATGTTTGTCTCAGCTATAA 1254
DB 1201 TCAAGTTAAGAGTCACATAATCCACAGGACATGTTTGTCTCAGCTTAGAA 1254
RESULT 5
LOCUS AR653694
DEFINITION Sequence 7 from patent US 6887696.
ACCESSION AR653694
VERSION AR653694.1 GI:67584279
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1278)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 7 03-MAY-2005;
LARGE SCALE Biology Corporation; Vacaville, CA
FEATURES
source
1..1278
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 99.9%; Score 1252.4; DB 6; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCGCTTCCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCGCTTCCTGGCC 60
QY 61 CTGCTTTCCTGGGACATCCCTGGGCTAGAGCACTGGACAAATGGATGGCAAGGACGCT 120
DB 61 CTGCTTTCCTGGGACATCCCTGGGCTAGAGCACTGGACAAATGGATGGCAAGGACGCT 120
QY 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
DB 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
QY 181 GATTCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAGGC 240
```



```
Db 721 AGTATCTTGGACTGCACATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTGGACAGGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAAATTTGGGCTCAGCTGGAAATCAGCAA 840
Db 781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAAATTTGGGCTCAGCTGGAAATCAGCAA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
QY 901 CCACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGCACTTAATGCCATCAAT 960
Db 901 CCACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGCACTTAATGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTAACCACTTTAGACAGGAGACAACTTTGAAAGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTAACCACTTTAGACAGGAGACAACTTTGAAAGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATATAAACCAGGAGATGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATATAAACCAGGAGATGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGAGGAGTGGCTGTAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGAGGAGTGGCTGTAATCCT 1140
QY 1141 GCCTGCTTCATCACACAGCTCTCCTGTGAAAGGAAGCTAGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCTCCTGTGAAAGGAAGCTAGGTTCTATGAATGGACT 1200
QY 1201 TCAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTAGAA 1254
Db 1201 TCAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTAGAA 1254

RESULT 7
AR653697
LOCUS 1284 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 13 from patent US 6887696.
ACCESSION AR653697
VERSION AR653697.1 GI:67584282
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1284)
AUTHORS Garger, S. J., Turpen, T. H. and Kumagai, M. H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 13 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
Location/Qualifiers
source 1..1284
/mol_type="genomic DNA"

ORIGIN
Query Match 99.9%; Score 1252.4; DB 6; Length 1284;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
QY 61 CTGCTTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGATTTGGCAAGCAAGCCCT 120
Db 61 CTGCTTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGATTTGGCAAGCAAGCCCT 120
QY 121 ACCATGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGATGCCAGGAGGCCA 180
Db 121 ACCATGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGATGCCAGGAGGCCA 180
QY 181 GATTCTTCGATCAGTGAGAGCTCTTCATGGAGATGCGAGAGCTCATGGTCTCAGAGGC 240
```

RESULT 8

```
181 GATTCTTCGATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
241 TGGAAAGGATCAGAGTTATGAGTACCTCTGCATTGATGACTGTTGGATGCTCCCAAGCA 300
241 TGGAAAGGATCAGAGTTATGAGTACCTCTGCATTGATGACTGTTGGATGCTCCCAAGCA 300
301 GATTTCAGAAAGGAGCAGACTTCAGGCAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
301 GATTTCAGAAAGGAGCAGACTTCAGGCAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
361 GCTAATATATGTTTACAGCAAAAGGAGCTGAAGCTAGGATTTATGAGATGTTGGAATAAA 420
361 GCTAATATGTTTACAGCAAAAGGAGCTGAAGCTAGGATTTATGAGATGTTGGAATAAA 420
421 ACCTGGCAGAGGCTTCCCTGGGAGTTTGGATACCTAGCGCTTTCCTCATGGGATTCGCCAGCTT 480
421 ACCTGGCAGAGGCTTCCCTGGGAGTTTGGATACCTAGCGATTTGATGATGCTGCAAGATTTG 480
481 GACTGGGAGTAGATCTGCTAAAAATTTGATGTTGTTACTGTGACAGATTTGGAATAATTG 540
481 GACTGGGAGTAGATCTGCTAAAAATTTGATGTTGTTACTGTGACAGATTTGGAATAATTG 540
541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 600
541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 600
601 TCCTGTGAGTGGCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
601 TCCTGTGAGTGGCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
661 CAGTACTGCAATCAGCTGGCGAAATTTGCTGCAATTTGATGATGATGATGATGATGATGATG 720
661 CAGTACTGCAATCAGCTGGCGAAATTTGCTGCAATTTGATGATGATGATGATGATGATGATG 720
721 AGTATCTTCGAGCTGGACATCTTTTAAACAGGAGAGAAATTTGATGATGTTGCTGACACAGG 780
721 AGTATCTTCGAGCTGGACATCTTTTAAACAGGAGAGAAATTTGATGATGTTGCTGACACAGG 780
781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCTCAGCTGGAAATCAGCAA 840
781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCTCAGCTGGAAATCAGCAA 840
841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGCACTTAATGCCATCAAT 960
901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGCACTTAATGCCATCAAT 960
961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAAGTGTGG 1020
961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAAGTGTGG 1020
1021 GAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATATAAACCAGGAGATGGT 1080
1021 GAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATATAAACCAGGAGATGGT 1080
1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGAGGAGTGGCTGTAATCCT 1140
1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGAGGAGTGGCTGTAATCCT 1140
1141 GCCTGCTTCATCACACAGCTCTCCTGTGAAAGGAAGCTAGGTTCTATGAATGGACT 1200
1141 GCCTGCTTCATCACACAGCTCTCCTGTGAAAGGAAGCTAGGTTCTATGAATGGACT 1200
1201 TCAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTAGAA 1254
1201 TCAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTAGAA 1254
```

Db	721	AGTATCTTGGACTGGGACATCTTTTAAACGAGGAGAAATTTGTGATGTTGCTGCACGAGG	780
Qy	781	GTTTGAATGACCCAGATATGTTAGTGATTTGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA	840
Db	781	GTTTGGATGACCCAGATATGTTAGTGATTTGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA	840
Qy	841	GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTAAATGACCTC	900
Db	841	GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTAAATGACCTC	900
Qy	901	CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAAGACGTAAATGGCATCAAT	960
Db	901	CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAAGACGTAAATGGCATCAAT	960
Qy	961	CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACCTTTGAAAGTGTGG	1020
Db	961	CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACCTTTGAAAGTGTGG	1020
Qy	1021	GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCGCAGGAGATTGGT	1080
Db	1021	GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCGCAGGAGATTGGT	1080
Qy	1081	GGACCTCGCTCTTATACCATCGAGTTGCTTCCTGGGTAAAGAGTGGCTGTATTCCT	1140
Db	1081	GGACCTCGCTCTTATACCATCGAGTTGCTTCCTGGGTAAAGAGTGGCTGTATTCCT	1140
Qy	1141	GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATCGACT	1200
Db	1141	GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATCGACT	1200
Qy	1201	TCAAGGTTAAGAGTCACATAATCCACAGGCACTGTTTTGCTTCAGCTATAA	1254
Db	1201	TCAAGGTTAAGAGTCACATAATCCACAGGCACTGTTTTGCTTCAGCTATAA	1254
RESULT 9			
CS135670		1290 bp	DNA
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
ORIGIN			
Query Match			
Best Local Similarity			
Matches 1253; Conservative			
Qy			
Db			
Qy			

Db 61 CTCGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGCAATGGATTGGCAAGGAGCGCT 120
QY 121 ACCATGGGCTGCTGACATGGGAGGCGCTTCATGTGCAACCTTTGACCTGCCAGGAAGGCCA 180
Db 121 ACCATGGGCTGCTGACATGGGAGGCGCTTCATGTGCAACCTTTGACCTGCCAGGAAGGCCA 180
QY 181 GATTCTCTGATCAGTGAGAGGCTCTTCATGGAGATGGCAGAGCTCATGTCTCTCAGAAGGC 240
Db 181 GATTCTCTGATCAGTGAGAGGCTCTTCATGGAGATGGCAGAGCTCATGTCTCTCAGAAGGC 240
QY 241 TCGAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTTGATGCTCCCAAGA 300
Db 241 TCGAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTTGATGCTCCCAAGA 300
QY 301 GATTCTGAGAGGAGCTTCAGGAGACCTCTAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTCTGAGAGGAGCTTCAGGAGACCTCTAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATTATGTTACAGCAAGGAGCTGAAGCTAGGATTTATGCAAGTCTTGGAAATAA 420
Db 361 GCTAATTATGTTACAGCAAGGAGCTGAAGCTAGGATTTATGCAAGTCTTGGAAATAA 420
QY 421 ACCTGGCAGGCTTCCTCGGAGTTTGGATACCTACGACATTCATGCCAGACCTTTGCT 480
Db 421 ACCTGGCAGGCTTCCTCGGAGTTTGGATACCTACGACATTCATGCCAGACCTTTGCT 480
QY 481 GACTGGGGAGTAGACTGCTTAAATTTGATGGTTGTA CTGTGACAGTTTGGAAATTTG 540
Db 481 GACTGGGGAGTAGACTGCTTAAATTTGATGGTTGTA CTGTGACAGTTTGGAAATTTG 540
QY 541 GCGATGCTTATGACACATGCTTCTGGCCCTGAATAGGACTGGCAGAGCAATGTTGTAC 600
Db 541 GCGATGCTTATGACACATGCTTCTGGCCCTGAATAGGACTGGCAGAGCAATGTTGTAC 600
QY 601 TCCGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTGCTGCAATTTGATGATTTCTGGAAAGATATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTGCTGCAATTTGATGATTTCTGGAAAGATATAAG 720
QY 721 AGTATCTTGACCTGGACATCTTTTAAACCAGGAGAAATTTGATGTTGTCGACACAGG 780
Db 721 AGTATCTTGACCTGGACATCTTTTAAACCAGGAGAAATTTGATGTTGTCGACACAGG 780
QY 781 GGTGGAAATGACCCAGATATGTTAGTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 841 GTAACCTCAGATGCCCCTCTGGGCTATCAGGCTGCTCCTTTATTCATGTCTAATGACCTC 900
Db 841 GTAACCTCAGATGCCCCTCTGGGCTATCAGGCTGCTCCTTTATTCATGTCTAATGACCTC 900
QY 901 CGACATCAGCCCTCAGGCAAGCTCTCCTTCAGGATAGGACGCTAATTTGGCCATCAAT 960
Db 901 CGACATCAGCCCTCAGGCAAGCTCTCCTTCAGGATAGGACGCTAATTTGGCCATCAAT 960
QY 961 CAGGACCCCTTGGCAGCAAGGATACAGCTTAGACAGGAGCAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGCAGCAAGGATACAGCTTAGACAGGAGCAACTTTGAAGTGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCAGGAGGATGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCAGGAGGATGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGGAGGCTGTAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGGAGGCTGTAATCCT 1140
QY 1141 GCCTGCTTTCATACACAGCTCTCCCTGTGAAAGGAGGCTAGGGTTCTATGATGGACT 1200
Db 1141 GCCTGCTTTCATACACAGCTCTCCCTGTGAAAGGAGGCTAGGGTTCTATGATGGACT 1200

QY 1201 TCAAGGTTAAGAGTCAATAATCCCAAGGCACTGTTTCTTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAGTCAATAATCCCAAGGCACTGTTTCTTCAGCTAGAA 1254
RESULT 10
LOCUS AR653692 1290 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 3 from patent US 6887696.
ACCESSION AR653692
VERSION AR653692.1 GI:67584277
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1290)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 3 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source 1..1290
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 99.9%; Score 1252.4; DB 6; Length 1290;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAGCTCAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTCGGCC 60
Db 1 ATGCAGCTCAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTCGGCC 60
QY 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGAGCGCT 120
Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGAGCGCT 120
QY 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTCAGCTGCCAGGAAGGCCA 180
Db 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTCAGCTGCCAGGAAGGCCA 180
QY 181 GATTCTCTGATCAGTGAGAGGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Db 181 GATTCTCTGATCAGTGAGAGGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
QY 241 TCGAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTTGATGCTCCCAAGA 300
Db 241 TCGAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTTGATGCTCCCAAGA 300
QY 301 GATTCTGAGAGGAGCTTCAGGAGACCTCTAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTCTGAGAGGAGCTTCAGGAGACCTCTAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATTATGTTACAGCAAGGAGCTGAAGCTAGGATTTATGCAAGTCTTGGAAATAA 420
Db 361 GCTAATTATGTTACAGCAAGGAGCTGAAGCTAGGATTTATGCAAGTCTTGGAAATAA 420
QY 421 ACCTGGCAGGCTTCCTCGGAGTTTGGATACCTACGACATTCATGCCAGACCTTTGCT 480
Db 421 ACCTGGCAGGCTTCCTCGGAGTTTGGATACCTACGACATTCATGCCAGACCTTTGCT 480
QY 481 GACTGGGGAGTAGACTGCTTAAATTTGATGGTTGTA CTGTGACAGTTTGGAAATTTG 540
Db 481 GACTGGGGAGTAGACTGCTTAAATTTGATGGTTGTA CTGTGACAGTTTGGAAATTTG 540
QY 541 GCGAGTGGTTTATAGCACATGTCTTCGGCCCTGAATAGGAGCTGGCAGAGCAATTTGTAC 600
Db 541 GCGAGTGGTTTATAGCACATGTCTTCGGCCCTGAATAGGAGCTGGCAGAGCAATTTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660

```
Db 601 TCCTGTGAGTGGCCCTCTTTATATATGTGGCCCTTTCAAAGGCCAATATATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCATCTGGCGAAATTTTGTCTGACATTTGATGATTCCTCGAAAGAGTATAAG 720
Db 661 CAGTACTGCAATCATCTGGCGAAATTTTGTCTGACATTTGATGATTCCTCGAAAGAGTATAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACGAGGAGAAATTTGATGATTCCTGGACGAGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACGAGGAGAAATTTGATGATTCCTGGACGAGG 780
Qy 781 GGTGGAAATGACCCAGAGATATGTTAGTATGGAATTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGAGATATGTTAGTATGGAATTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACATCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCCTAATGACCTC 900
Db 841 GTAACATCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCCTAATGACCTC 900
Qy 901 CGACATCATGACCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATGGCATCAAT 960
Db 901 CGACATCATGACCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATGGCATCAAT 960
Qy 961 CAGGACCCCTTTGGCAAGCAAGGTACAGCTTAGACAGGAGACAACTTTTGAAGTGTG 1020
Db 961 CAGGACCCCTTTGGCAAGCAAGGTACAGCTTAGACAGGAGACAACTTTTGAAGTGTG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTTAGCCTGGGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTTAGCCTGGGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080
Qy 1081 GGAAGCTGCTTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGTGGCTGTAAATCCT 1140
Db 1081 GGAAGCTGCTTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGTGGCTGTAAATCCT 1140
Qy 1141 GCCTGCTTCAATCACAGCTCTCTCCCTGTGAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCAATCACAGCTCTCTCCCTGTGAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTTGCTTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTTGCTTCAGCTATGAA 1254

RESULT 11
AR654960
LOCUS AR654960 1290 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 3 from patent US 6890748.
ACCESSION AR654960
VERSION AR654960.1 GI:67586497
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1290)
AUTHORS Garger, S.J., Turpen, T.H. and Kunagai, M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 3 10-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source Location/Qualifiers
1..1290
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 99.98; Score 1252.4; DB 6; Length 1290;
Best Local Similarity 99.98; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCCGTTCCCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGGCTTCCGTTCCCTGGCC 60
Qy 61 CTCGTTTCCTGGGACATCCCTGGGGCTAGAGCACTGGACAATGATTTGGCAAGGACGCT 1200
```

```
Db 61 CTCGTTTCCTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGACGCT 120
Qy 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
Db 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
Qy 181 GATTCTTCATAGTGAAGAGCTCTTCATGAGAGTGGCAGAGCTCATGTGTCAGAGAGG 240
Db 181 GATTCTTCATAGTGAAGAGCTCTTCATGAGAGTGGCAGAGCTCATGTGTCAGAGAGG 240
Qy 241 TGGAGGATCGAGTTATGAGTACCTCTGATGATGACCTGTTGGATGGCTCCCAAGA 300
Db 241 TGGAGGATCGAGTTATGAGTACCTCTGATGATGACCTGTTGGATGGCTCCCAAGA 300
Qy 301 GATTTCAGAGAGCAGACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCGAGCTA 360
Db 301 GATTTCAGAGAGCAGACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCGAGCTA 360
Qy 361 GCTAATATGTTTCAAGCAAGGACTGAACTAGGGAATTTATGAGATGTTGGAATAA 420
Db 361 GCTAATATGTTTCAAGCAAGGACTGAACTAGGGAATTTATGAGATGTTGGAATAA 420
Qy 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACCTAGACATGATGCCAGACCTTTGCT 480
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACCTAGACATGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGGAGTAGACTGCTTAAATTTGATGGTTGTTTACTGTGACAGTTTGGAAAAATTTG 540
Db 481 GACTGGGGAGTAGACTGCTTAAATTTGATGGTTGTTTACTGTGACAGTTTGGAAAAATTTG 540
Qy 541 GCAGATGTTTAAAGCAATGTCCTTGGCCCTGAAATAGGACTGGCAGAGAGATTGTGTAC 600
Db 541 GCAGATGTTTAAAGCAATGTCCTTGGCCCTGAAATAGGACTGGCAGAGAGATTGTGTAC 600
Qy 601 TCCTGTGAGTGGCTCTTTTATATGTCGCCCTTTCAAAGGCCAATATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTTATATGTCGCCCTTTCAAAGGCCAATATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCATCTGGCGAAATTTTGTCTGACATTTGATGATTCCTCGAAAGAGTATAAG 720
Db 661 CAGTACTGCAATCATCTGGCGAAATTTTGTCTGACATTTGATGATTCCTCGAAAGAGTATAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACGAGGAGAAATTTGATGATTCCTGGACGAGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACGAGGAGAAATTTGATGATTCCTGGACGAGG 780
Qy 781 GGTGGAAATGACCCAGAGATATGTTAGTATGGAATTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGAGATATGTTAGTATGGAATTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACATCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCCTAATGACCTC 900
Db 841 GTAACATCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCCTAATGACCTC 900
Qy 901 CGACATCATGACCCCTCAAGCAAGGCTCTCTTCAGGATAGGACGTAATGGCATCAAT 960
Db 901 CGACATCATGACCCCTCAAGCAAGGCTCTCTTCAGGATAGGACGTAATGGCATCAAT 960
Qy 961 CAGGACCCCTTTGGCAAGCAAGGTACAGCTTAGACAGGAGACAACTTTTGAAGTGTG 1020
Db 961 CAGGACCCCTTTGGCAAGCAAGGTACAGCTTAGACAGGAGACAACTTTTGAAGTGTG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTTAGCCTGGGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTTAGCCTGGGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080
Qy 1081 GGAAGCTGCTTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGTGGCTGTAAATCCT 1140
Db 1081 GGAAGCTGCTTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGTGGCTGTAAATCCT 1140
Qy 1141 GCCTGCTTCAATCACAGCTCTCTCCCTGTGAAGGAAGCTAGGGTTCTATGAATGGACT 1200
```


Db 1081 GGACCTCGCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCCTGTATCCT 1140

Qy 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGAGAGCTAGGGTTCTATGAATGGACT 1200

Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGAGAGCTAGGGTTCTATGAATGGACT 1200

Qy 1201 TCAAGGTTAAGAGTACATATAATCCACAGGCACTGTTTGGCTTCAGCTATAA 1254

Db 1201 TCAAGGTTAAGAGTACATATAATCCACAGGCACTGTTTGGCTTCAGCTATAA 1254

RESULT 13

AY889631

LOCUS

DEFINITION

AY889631 1290 bp mRNA linear SYN 29-MAR-2005

alpha (GLA) mRNA, complete cds.

ACCESSION

AY889631

VERSION

AY889631.1 GI:60655716

KEYWORDS

Human ORF Project.

SOURCE

synthetic construct

ORGANISM

other sequences; artificial sequences.

REFERENCE

1 (bases 1 to 1290)

AUTHORS

Hines, L., Rolfs, A., Jepson, D., Moreira, D., Raphael, J., Kelley, F., Shen, B., Halleck, A., Koundinya, M., Hu, Y., Zuo, D., Taycher, E., Williamson, J. and LaBaer, J.

TITLE

Cloning of human full-length CDS in Creator (TM) recombinational vector system

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 1290)

AUTHORS

Hines, L., Rolfs, A., Jepson, D., Moreira, D., Raphael, J., Kelley, F., Shen, B., Halleck, A., Koundinya, M., Hu, Y., Zuo, D., Taycher, E., Williamson, J. and LaBaer, J.

TITLE

Direct Submission

JOURNAL

Submitted (04-JAN-2005) Biological Chemistry and Molecular Pharmacology, Harvard Institute of Proteomics, 320 Charles St., Cambridge, MA 02141, USA

COMMENT

This CDS clone is a part of a collection of human full-length expression clones generated by Harvard Institute of Proteomics. This ORF clone has been cloned with normalized stop-codon. The CDS has been directionally cloned using BD In-Fusion(TM) cloning system between the SalI and HindIII sites of the pDNR-Dual vector. Additional sequences in the clone: 'ACC' after SalI site and before 'ATG' to provide Kozak consensus sequence. Each clone is clonally isolated and full-length sequence-verified.

FEATURES

Source

1..1290

/organism="synthetic construct"

/mol_type="mRNA"

/db_xref="taxon:32630"

/clone="FLH025769.01X"

/lab_host="Escherichia coli DH5alpha T1 resistant"

/note="derived from MGC template"

gene

1..1290

/genes="GLA"

CDS

1..1290

/genes="GLA"

/codon_start=1

/transl_table=11

/product="galactosidase alpha"

/protein_id="AA32422.1"

/db_xref="GI:60655717"

/translation="MOLRNPGLHGCALALRFLALVSWDIPGARALDNLGARTPTMGW LHERPWCNLDQBERPDCISEKLFMEAEALVSEGMKDGAYEYLICIDCWMAPOKDS EGRLOAPQRPFGHIGRLANYVHSGKLGIYADVGNKTCAGFPSPGFGYDIDAQTPA DNGVDLTKFDGCTCDSENLADGKRLMSALARTGRSIVTSCWEPDLMPPFPKNTTE IRQYCNHNRNADIDDSWKSISLDWTSFNQERIVDVAGPGGNDPDLVIGNFGLS WNOQVQMALWAIMAPLFWMSNDRHISPAKALLQDKVIAINQDPLGKQGYLRFQG DNPVWFRPLSLAWAVAMINROEIGGPRSYTAVASLGKGVACNPACFITQLLPVKR LKGFYEWTSLRSHINPTGTVLLQLENTWQMSLKDL"

ORIGIN

99.9%; Score 1252.4; DB 11; Length 1290;

Query Match

Best Local Similarity 99.9%; Pred. No. 0; Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCTCGTGGCC 60

Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCTCGTGGCC 60

Qy 61 CTGCTTCTTGGGACATCCCTGGGCTTAGAGCACTGGACAATGGATTGGCAAGACGCT 120

Db 61 CTGCTTCTTGGGACATCCCTGGGCTTAGAGCACTGGACAATGGATTGGCAAGACGCT 120

Qy 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGGCTTCTCGGAGAGCCA 180

Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGGCTTCTCGGAGAGCCA 180

Qy 181 GATTCTCTGCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAGGC 240

Db 181 GATTCTCTGCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAGGC 240

Qy 241 TGAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTGGATGGCTCCCAAGA 300

Db 241 TGAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTGGATGGCTCCCAAGA 300

Qy 301 GATTCAAGAGGCAGACTTCAGGAGAGCCCTTCAGGCTTTCCTCATGGATTCGCGACTA 360

Db 301 GATTCAAGAGGCAGACTTCAGGAGAGCCCTTCAGGCTTTCCTCATGGATTCGCGACTA 360

Qy 361 GCTAATTTATGTTTCAACAGCAAGAGCTGAAGCTAGGAGATTTATGCAGATGTTGGAATAA 420

Db 361 GCTAATTTATGTTTCAACAGCAAGAGCTGAAGCTAGGAGATTTATGCAGATGTTGGAATAA 420

Qy 421 ACCTGCGCAGGCTTCCCTGGGAGTTTTGGATACCTACGACATTCATGATGCTTGTGCT 480

Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTTGGATACCTACGACATTCATGATGCTTGTGCT 480

Qy 481 GACTGGGAGTAGATCTGCTTAAATTTGATGTTGTTTACTGTGACAGTTTGGAAATTTG 540

Db 481 GACTGGGAGTAGATCTGCTTAAATTTGATGTTGTTTACTGTGACAGTTTGGAAATTTG 540

Qy 541 GCAGATGGTTTATAAGCACATGCTCTGGGCTGCTGAATAGGACTGGCAGACATTTGTGCTAC 600

Db 541 GCAGATGGTTTATAAGCACATGCTCTGGGCTGCTGAATAGGACTGGCAGACATTTGTGCTAC 600

Qy 601 TCTGTGAGTGGCTCTTTATATGTCGCTTTCAAAAGGCCCAATATATACAGAAATCCGA 660

Db 601 TCTGTGAGTGGCTCTTTATATGTCGCTTTCAAAAGGCCCAATATATACAGAAATCCGA 660

Qy 661 CAGTACTGCAATCAGTGGGCAATTTTGTGCAATTCATGATGATTCCTGGAAAGTATAAG 720

Db 661 CAGTACTGCAATCAGTGGGCAATTTTGTGCAATTCATGATGATTCCTGGAAAGTATAAG 720

Qy 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACCAAGG 780

Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACCAAGG 780

Qy 781 GGTGGAATGACCCAGATATGTTAGTATGTTGGCACTTTGGCTTCAGCTGGAATCAGCAA 840

Db 781 GGTGGAATGACCCAGATATGTTAGTATGTTGGCACTTTGGCTTCAGCTGGAATCAGCAA 840

Qy 841 GTAATCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900

Db 841 GTAATCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900

Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGCTAAATGGCCATCAAT 960

Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGCTAAATGGCCATCAAT 960

Qy 961 CAGGACCCCTTGGGCAAGCAAGGCTACCGAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1020

Db 961 CAGGACCCCTTGGGCAAGCAAGGCTACCGAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1020

Qy 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATATAACCGGAGGAGATTGGT 1080

Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATATAACCGGAGGAGATTGGT 1080

Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATATAACCGGCGAGAGATTGCT 1080
QY 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGAGAGTGGCTGTATATCTT 1140
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGAGAGTGGCTGTATATCTT 1140
QY 1141 GCCTGCTTATCACACAGCTCTCTCCCTGTGAAAGGAAGCTAGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTATCACACAGCTCTCTCCCTGTGAAAGGAAGCTAGGTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTACATATAATCCACAGGCACTGTTTGTCTTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAGTACATATAATCCACAGGCACTGTTTGTCTTCAGCTAGTA 1254

RESULT 14
AY889632
LOCUS
DEFINITION
1290 bp mRNA linear SYN 29-MAR-2005
Synthetic construct Homo sapiens clone FLH025770.01X galactosidase
alpha (GLA) mRNA, complete cds.
ACCESSION
AY889632
VERSION
AY889632.1 GI:60655718
KEYWORDS
Human ORF Project.
SOURCE
synthetic construct
ORGANISM
synthetic construct
other sequences; artificial sequences.
REFERENCE
1 (bases 1 to 1290)
Hines, L., Rolfs, A., Jepson, D., Moreira, D., Raphael, J., Kelley, F.,
Shen, B., Halleck, A., Koundinya, M., Hu, Y., Zuo, D., Taycher, E.,
Williamson, J. and Labaer, J.
AUTHORS
Cloning of human full-length CDS in Creator (TM) recombinational
vector system
Unpublished
REFERENCE
2 (bases 1 to 1290)
Hines, L., Rolfs, A., Jepson, D., Moreira, D., Raphael, J., Kelley, F.,
Shen, B., Halleck, A., Koundinya, M., Hu, Y., Zuo, D., Taycher, E.,
Williamson, J. and Labaer, J.
AUTHORS
Direct Submission
TITLE
Submitted (04-JAN-2005) Biological Chemistry and Molecular
Pharmacology, Harvard Institute of Proteomics, 320 Charles St.,
Cambridge, MA 02141, USA

This CDS clone is a part of a collection of human full-length
expression clones generated by Harvard Institute of Proteomics. CDS
has been directionally cloned using BD In-Fusion(TM) cloning system
between the SalI and HindIII sites of the pDNR-Dual vector.
Additional sequences in the clone: 'ACC' after SalI site and
before 'ATG' to provide Kozak consensus sequence. Each clone is
clonally isolated and full-length sequence-verified.
LOCATION/Qualifiers
1. .1290
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clone="FLH025770.01X"
/lab_host="Escherichia coli DH5alpha T1 resistant"
/note="derived from MGC template"
1. .1290
/gene="GLA"
1. .1290
/gene="GLA"
/codon_start=1
/transl_table=11
/product="galactosidase alpha"
/protein_id="AA32423.1"
/db_xref="GI:60655719"
/translation="MQLRNPDLHLGALALRFLALVSWDIPGARLDNGLARTPTMGW
LHWERWCNLDCEEEDSCISEKLFMEELMWSEGWKDCAGFYLYCIDDDCWAPQDPS
EGLQADPQRFPHGIQRLANYHSHKGLGIYADVGNKTCAGFPFGFYDYIDAQTPA
DWGDLIKFDCYCDSENLADYKIMSLALNTRGSIVYSEWPLMFMFPKPNYTE
IROYCNHRNFADIDSWKSIKSLDWTFSNOERIYDVAFPGWGNPDMLVIGNFGLS
WNOQVTQMALWATMAAPLWMSNDRHISPOKALLODKVIAINQDPLKGQYQLRGR
DNFEVWRPLSGLAWAVAMINRQIEIGGPRSYTIAVSLGKGVACNPAFCITQLLPVKR

FEATURES
source
1. .1290
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clone="FLH025770.01X"
/lab_host="Escherichia coli DH5alpha T1 resistant"
/note="derived from MGC template"
1. .1290
/gene="GLA"
1. .1290
/gene="GLA"
/codon_start=1
/transl_table=11
/product="galactosidase alpha"
/protein_id="AA32423.1"
/db_xref="GI:60655719"
/translation="MQLRNPDLHLGALALRFLALVSWDIPGARLDNGLARTPTMGW
LHWERWCNLDCEEEDSCISEKLFMEELMWSEGWKDCAGFYLYCIDDDCWAPQDPS
EGLQADPQRFPHGIQRLANYHSHKGLGIYADVGNKTCAGFPFGFYDYIDAQTPA
DWGDLIKFDCYCDSENLADYKIMSLALNTRGSIVYSEWPLMFMFPKPNYTE
IROYCNHRNFADIDSWKSIKSLDWTFSNOERIYDVAFPGWGNPDMLVIGNFGLS
WNOQVTQMALWATMAAPLWMSNDRHISPOKALLODKVIAINQDPLKGQYQLRGR
DNFEVWRPLSGLAWAVAMINRQIEIGGPRSYTIAVSLGKGVACNPAFCITQLLPVKR

gene
CDS
1. .1290
/gene="GLA"
1. .1290
/gene="GLA"
/codon_start=1
/transl_table=11
/product="galactosidase alpha"
/protein_id="AA32423.1"
/db_xref="GI:60655719"
/translation="MQLRNPDLHLGALALRFLALVSWDIPGARLDNGLARTPTMGW
LHWERWCNLDCEEEDSCISEKLFMEELMWSEGWKDCAGFYLYCIDDDCWAPQDPS
EGLQADPQRFPHGIQRLANYHSHKGLGIYADVGNKTCAGFPFGFYDYIDAQTPA
DWGDLIKFDCYCDSENLADYKIMSLALNTRGSIVYSEWPLMFMFPKPNYTE
IROYCNHRNFADIDSWKSIKSLDWTFSNOERIYDVAFPGWGNPDMLVIGNFGLS
WNOQVTQMALWATMAAPLWMSNDRHISPOKALLODKVIAINQDPLKGQYQLRGR
DNFEVWRPLSGLAWAVAMINRQIEIGGPRSYTIAVSLGKGVACNPAFCITQLLPVKR

ORIGIN

Query Match 99.9%; Score 1252.4; DB 11; Length 1290;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCGAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTGGCTTCCCTGGGCC 60
Db 1 ATGCGAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTGGCTTCCCTGGGCC 60
QY 61 CTCGTTTCTCTGGGACATCCCTCGGGCTAGAGCACTGGACAATGATTTGGCAAGACGCTT 120
Db 61 CTCGTTTCTCTGGGACATCCCTCGGGCTAGAGCACTGGACAATGATTTGGCAAGACGCTT 120
QY 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACTTTGACTGCCAGGAAGAGCCA 180
Db 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACTTTGACTGCCAGGAAGAGCCA 180
QY 181 GATTCCTGTCATCAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 181 GATTCCTGTCATCAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 241 TGGAGGATGCGAGGTTATGAGTACCTCTGCAATTTGATGACTTGTGGATGGCTCCCCAAAGA 300
Db 241 TGGAGGATGCGAGGTTATGAGTACCTCTGCAATTTGATGACTTGTGGATGGCTCCCCAAAGA 300
QY 301 GATTCAGAGGCGAGCTTCAGGCAAGACCTTCAGGAGATGGCAGAGCTCATGGTCTCAGGAGTAA 360
Db 301 GATTCAGAGGCGAGCTTCAGGCAAGACCTTCAGGAGATGGCAGAGCTCATGGTCTCAGGAGTAA 360
QY 361 GCTTAATTTGTTTACAGCAAGAAAGGACTGAAGCTAGGATTTATGACAGATGTTGAAATATAA 420
Db 361 GCTTAATTTGTTTACAGCAAGAAAGGACTGAAGCTAGGATTTATGACAGATGTTGAAATATAA 420
QY 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTTGATGCCAGACCTTTGCT 480
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTTGATGCCAGACCTTTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTTACTGTGACAGATTTGGAAATTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTTACTGTGACAGATTTGGAAATTTG 540
QY 541 GCAGATGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCACTTGTGTAC 600
Db 541 GCAGATGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCACTTGTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTTATATGTCGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTTATATGTCGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGGAAATTTTGTGACATTTGATGATTTCTCGGAAAGATATAAG 720
Db 661 CAGTACTGCAATCACTGGCGGAAATTTTGTGACATTTGATGATTTCTCGGAAAGATATAAG 720
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGATGTTTGTGACAGAGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGATGTTTGTGACAGAGG 780
QY 781 GGTGGAAATGACCCAGATATGTTAGTGATTTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTGATTTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
QY 841 GTAACCTAGATGGCCCTTGGGCTATCATGGCTGCTCTTTTATTCATGTTCTTAATGACCTC 900
Db 841 GTAACCTAGATGGCCCTTGGGCTATCATGGCTGCTCTTTTATTCATGTTCTTAATGACCTC 900
QY 901 CGACATCAGCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Db 901 CGACATCAGCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGATACCACTGTAGACAGGAGCAAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGATACCACTGTAGACAGGAGCAAACTTTGAAGTGTGG 1020

Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTGCCATCAAT 960
Db |||||
Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTGCCATCAAT 960
Db |||||
Qy 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
Db |||||
Qy 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
Db |||||
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGCAGGAGATTGGT 1080
Db |||||
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGCAGGAGATTGGT 1080
Db |||||
Qy 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
Db |||||
Qy 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
Db |||||
Qy 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Db |||||
Qy 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Db |||||
Qy 1201 TCAAGGTTAAGAAAGTACATAAATCCACAGGCACTGTTTGGCTTCAGCTATAA 1254
Db |||||
Qy 1201 TCAAGGTTAAGAAAGTACATAAATCCACAGGCACTGTTTGGCTTCAGCTAGAA 1254
Db |||||

Search completed: December 31, 2005, 19:26:44
Job time : 6626 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 31, 2005, 09:31:46 ; Search time 843 Seconds
(without alignments)
9914.027 Million cell updates/sec

Title: US-10-602-220-15
Perfect score: 1254
Sequence: 1. atgcagctgagggaaccaga.....ctgttttgcctgcagctataa 1254

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*
14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1254	100.0	1254	6	AAD45225 Human rga
2	1254	100.0	1254	10	ADD84754 Human rga
3	1254	100.0	1254	12	ADM48684 Human wil
4	1254	100.0	1254	13	ADU66919 Human alp
5	1254	100.0	1254	14	AEA27448 Human alp
6	1252.4	99.9	1266	6	AAD45223 Human rga
7	1252.4	99.9	1266	10	ADD84750 Human alp
8	1252.4	99.9	1266	12	ADJ88276 Human wt
9	1252.4	99.9	1266	12	ADM48680 Human wil
10	1252.4	99.9	1266	13	ADU66915 Human alp
11	1252.4	99.9	1266	14	AEA27444 Human alp
12	1252.4	99.9	1278	6	AAD45221 Human rga
13	1252.4	99.9	1278	10	ADD84746 Human alp
14	1252.4	99.9	1278	12	ADJ88272 Human rga
15	1252.4	99.9	1278	12	ADM48676 Human wil
16	1252.4	99.9	1278	13	ADU66911 Human alp
17	1252.4	99.9	1278	14	AEA27440 Human alp
18	1252.4	99.9	1284	6	AAD45224 Human rga
19	1252.4	99.9	1284	10	ADD84752 Human alp

20	1252.4	99.9	1284	12	ADJ88278 Human wt
21	1252.4	99.9	1284	12	ADM48682 Human wil
22	1252.4	99.9	1284	13	ADU66917 Human alp
23	1252.4	99.9	1284	14	AEA27446 Human alp
24	1252.4	99.9	1290	6	AAD45219 Human wt
25	1252.4	99.9	1290	10	ADD84742 Human alp
26	1252.4	99.9	1290	12	ADJ88268 Human wt
27	1252.4	99.9	1290	12	ADM48672 Human wil
28	1252.4	99.9	1290	13	ADU49739 Human alp
29	1252.4	99.9	1290	13	ADU74415 Human alp
30	1252.4	99.9	1290	13	ADU66907 Human alp
31	1252.4	99.9	1290	14	AEA27436 Human alp
32	1252.4	99.9	1290	14	AEA27436 Human alp
33	1252.4	99.9	1296	6	AAD45222 Human rga
34	1252.4	99.9	1296	10	ADD84748 Human alp
35	1252.4	99.9	1296	12	ADJ88274 Human wt
36	1252.4	99.9	1296	12	ADM48678 Human wil
37	1252.4	99.9	1296	13	ADU66913 Human alp
38	1252.4	99.9	1296	14	AEA27442 Human alp
39	1252.4	99.9	1306	8	ABZ79777 Human alp
40	1252.4	99.9	1308	6	AAD45220 Human wt
41	1252.4	99.9	1308	10	ADD84744 Human alp
42	1252.4	99.9	1308	12	ADJ88270 Human wt
43	1252.4	99.9	1308	12	ADM48674 Human wil
44	1252.4	99.9	1308	13	ADU66909 Human alp
45	1252.4	99.9	1308	14	AEA27438 Human alp

ALIGNMENTS

RESULT 1
AAD45225
ID AAD45225 standard; DNA; 1254 BP.
XX AAD45225;
XX
DT 27-DEC-2002 (first entry)
XX Human rGAL-12 DNA.
XX
KW Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
KW therapeutic; rGAL-12; gene; ds.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1254
FT /*tag= a
FT /product= "Human rGAL-12 protein"
XX
XX US2002088024-A1.
XX
XX 04-JUL-2002.
XX
XX 13-NOV-2001; 2001US-00993059.
XX
XX 26-JUL-2000; 2000US-00626127.
XX
XX (GARG/) GARGER S J.
XX (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2002-681656/73.
XX P-PSDB; AAE28212.
XX Novel human alpha-galactosidase polypeptide useful for treating lysosomal
XX storage diseases.
XX Claim 1; Page 47-49; 88pp; English.

CC The invention relates to human alpha-galactosidase truncated at the
CC carboxy terminus and the production of enzymatically active recombinant
CC human and animal lysosomal enzymes. The invention is useful for producing
CC lysosomal enzymes for treating lysosomal storage diseases, producing
CC altered or mutated proteins, enzymatically active or otherwise, to serve
CC as precursors or substrates for further in vivo or in vitro processing to
CC a specialised industrial form for research or therapeutic uses, to
CC produce more effective therapeutic enzyme, for producing antibodies
CC against lysosomal enzymes for medical diagnostic use, and in any
CC commercial process that involves substrate hydrolysis. The present
CC sequence is human rGAL-12 DNA
XX

SQ Sequence 1254 BP; 316 A; 289 C; 328 G; 321 T; 0 U; 0 Other;

Query Match 100.0%; Score 1254; DB 6; Length 1254;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGNACCAGCACTACATCTGGGCTGGCGCTTGGCTTCCTCGTGGCC 60

DB 1 ATGCAGCTGAGNACCAGCACTACATCTGGGCTGGCGCTTGGCTTCCTCGTGGCC 60

QY 61 CTCGTTTCTTGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGAGCCCT 120

DB 61 CTCGTTTCTTGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGAGCCCT 120

QY 121 ACCATGGGTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGGCCA 180

DB 121 ACCATGGGTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGGCCA 180

QY 181 GATTCTGTGATCAGTGAAGAGCTTCATGGAGATGGCAGAGCTCATGTCTCAGAGGC 240

DB 181 GATTCTGTGATCAGTGAAGAGCTTCATGGAGATGGCAGAGCTCATGTCTCAGAGGC 240

QY 241 TGAAGAGTGCAGGTTATGAGTACCTCTGATGATGAGCTGTGGATGGCTCCCAAGA 300

DB 241 TGAAGAGTGCAGGTTATGAGTACCTCTGATGATGAGCTGTGGATGGCTCCCAAGA 300

QY 301 GATTTCAGAGGCGAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGATTCGCCAGCTA 360

DB 301 GATTTCAGAGGCGAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGATTCGCCAGCTA 360

QY 361 GCTAATATGTTTACAGCAAGAGCTGAGCTAGGAGTTTATGACATGTTGGAAATAA 420

DB 361 GCTAATATGTTTACAGCAAGAGCTGAGCTAGGAGTTTATGACATGTTGGAAATAA 420

QY 421 ACCTGCGCAGGCTTCCTGGGAGTTTGGATACCTAGCATTGATGAGCTGATGAGCTTTGCT 480

DB 421 ACCTGCGCAGGCTTCCTGGGAGTTTGGATACCTAGCATTGATGAGCTGATGAGCTTTGCT 480

QY 481 GACTGGGGAGTAGATCTGCTAAAATTGATGGTTGTTACTGTGACAGATTTGGAAAATTG 540

DB 481 GACTGGGGAGTAGATCTGCTAAAATTGATGGTTGTTACTGTGACAGATTTGGAAAATTG 540

QY 541 GCAGATGTTTAAAGCAGATCTCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600

DB 541 GCAGATGTTTAAAGCAGATCTCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600

QY 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660

DB 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660

QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTTGGAAAAGTATAAG 720

DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTTGGAAAAGTATAAG 720

QY 721 AGTATCTTGACATGGAATTTTAAACAGAGAGATTTGATGATTTGTCGACAGG 780

DB 721 AGTATCTTGACATGGAATTTTAAACAGAGAGATTTGATGATTTGTCGACAGG 780

QY 781 GGTGTGAATCACCAGATATGTTAGTATGTTGGAACTTTGGCCCTCAGCTGGAATCAGCAA 840

DB 781 GGTGTGAATCACCAGATATGTTAGTATGTTGGAACTTTGGCCCTCAGCTGGAATCAGCAA 840

QY 841 GTAACTCAGATGGCCTCTGGGCTATCATGGCTGCTCCTTTATTTCATGTCATGACCTC 900

DB 841 GTAACTCAGATGGCCTCTGGGCTATCATGGCTGCTCCTTTATTTCATGTCATGACCTC 900

QY 901 CGACACATCAGCCCTCAAGCCAAAAGCTCTCCTTCAGGATAAGGACGTAATTCATCAAT 960

DB 901 CGACACATCAGCCCTCAAGCCAAAAGCTCTCCTTCAGGATAAGGACGTAATTCATCAAT 960

QY 961 CAGGACCCCTTGGGCAAGCAAGGGTACAGCTTACAGCTTAGACAGGGAGACAACTTTGAAGTGG 1020

DB 961 CAGGACCCCTTGGGCAAGCAAGGGTACAGCTTACAGCTTAGACAGGGAGACAACTTTGAAGTGG 1020

QY 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080

DB 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080

QY 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTTAAATCCT 1140

DB 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTTAAATCCT 1140

QY 1141 GCCTGCTTATCATCAGAGCTCCTCCTGTGAAAAGGAGCTAGGTTCTATGATGAGT 1200

DB 1141 GCCTGCTTATCATCAGAGCTCCTCCTGTGAAAAGGAGCTAGGTTCTATGATGAGT 1200

QY 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGACCTGTTTGTCTCAGCTATAA 1254

DB 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGACCTGTTTGTCTCAGCTATAA 1254

RESULT 2

ADD84754

ID ADD84754 standard; DNA; 1254 BP.

XX ADD84754;

XX AC AC

XX 29-JAN-2004 (first entry)

XX Human alpha-galactosidase rGAL-12 DNA.

XX Human; alpha-galactosidase; rGAL-4; gene; ds; lysosomal enzyme;

XX enzyme replacement therapy; lysosomal disease.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 1..1254

XX /*tag= a

XX /product= "Human rGAL-12"

XX US2003106095-A1.

XX 05-JUN-2003.

XX 20-MAR-2002; 2002US-00103327.

XX 26-JUL-2000; 2000US-00626127.

XX 13-NOV-2001; 2001US-00993059.

XX (GARG/) GARGER S J.

XX (TURP/) TURPEN T H.

XX (KUMA/) KUMAGAI M H.

XX Garger SJ, Turpen TH, Kumagai MH;

XX WPI; 2003-801257/75.

XX P-PSDB; ADD84755.

XX New polynucleotide for producing active recombinant human and animal

XX lysosomal enzymes in a plant expression system that can be used in enzyme

XX replacement therapy.

XX Claim 1; SEQ ID NO 15; 77pp; English.

XX The invention relates to human alpha-galactosidase derivatives and the
CC nucleic acids encoding them. The polypeptides are used in a method for
CC producing active recombinant human and animal lysosomal enzymes in a
CC plant expression system. The enzymes can be used in enzyme replacement
CC therapy for the therapeutic treatment of human and animal lysosomal
CC diseases. This sequence represents DNA encoding a human alpha-
XX galactosidase derivative polypeptide of the invention.

SQ Sequence 1254 BP; 316 A; 289 C; 328 G; 321 T; 0 U; 0 Other;

Query Match 100.0%; Score 1254; DB 10; Length 1254;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCCTGGCC 60
DB 1 ATCCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCCTGGCC 60

QY 61 CTCGTTTCTGGGACATCTCTGGGCTAGAGCACTGGCAATGATGGCAAGGACGCT 120
DB 61 CTCGTTTCTGGGACATCTCTGGGCTAGAGCACTGGCAATGATGGCAAGGACGCT 120

QY 121 ACCATGGGCTGGCTGACCTGGGCGCTTCTATGCAACCTTCACTGCCAGGAGAGCCA 180
DB 121 ACCATGGGCTGGCTGACCTGGGCGCTTCTATGCAACCTTCACTGCCAGGAGAGCCA 180

QY 181 GATTCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
DB 181 GATTCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

QY 241 TGGAGGATGACGGTTATGATGATACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
DB 241 TGGAGGATGACGGTTATGATGATACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300

QY 301 GATTGAGAGGACACTTCAGGAGACCCCTCAGGCTTCTCATGGGATTCGCCAGCTA 360
DB 301 GATTGAGAGGACACTTCAGGAGACCCCTCAGGCTTCTCATGGGATTCGCCAGCTA 360

QY 361 GCTAAATATGTTTACAGCAAGGACTGAGCTAGGAGATTTATGAGATGTTGGAATAAA 420
DB 361 GCTAAATATGTTTACAGCAAGGACTGAGCTAGGAGATTTATGAGATGTTGGAATAAA 420

QY 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATGATGCCAGACCTTTGCT 480
DB 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATGATGCCAGACCTTTGCT 480

QY 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
DB 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540

QY 541 GCAGATGGTTTATAAGCACATGCTCTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC 600
DB 541 GCAGATGGTTTATAAGCACATGCTCTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC 600

QY 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660

QY 661 CAGTACTGCATCAGCTGGCGAAATTTTGTGACATTTGATGATGATTCCTGGAAAGTATAAG 720
DB 661 CAGTACTGCATCAGCTGGCGAAATTTTGTGACATTTGATGATGATTCCTGGAAAGTATAAG 720

QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGACACAGGG 780
DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGACACAGGG 780

QY 781 GGTGGAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
DB 781 GGTGGAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840

QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTAATGACCTC 900
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTAATGACCTC 900

DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTAATGACCTC 900

QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT 960

DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT 960

QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACCAAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1020

DB 961 CAGGACCCCTTGGGCAAGCAAGGTTACCAAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1020

QY 1021 GAACGACCTCTCTCAGGCTTTAGCTCGGCTGTAGCTATGATAAACCGGACGAGAGATTGGT 1080

DB 1021 GAACGACCTCTCTCAGGCTTTAGCTCGGCTGTAGCTATGATAAACCGGACGAGAGATTGGT 1080

QY 1081 GGACCTCGCTCTTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTCTGAATCCT 1140

DB 1081 GGACCTCGCTCTTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTCTGAATCCT 1140

QY 1141 GCCTGCTTTCATCACAGCTCTCTCCCTGTGAAAGGAAGCTAGGTTCTATGATGGACT 1200

DB 1141 GCCTGCTTTCATCACAGCTCTCTCCCTGTGAAAGGAAGCTAGGTTCTATGATGGACT 1200

QY 1201 TCAAGGTTAAGAGTGCACATAAATCCACAGGCACCTGTTTGTCTCAGCTATAA 1254

DB 1201 TCAAGGTTAAGAGTGCACATAAATCCACAGGCACCTGTTTGTCTCAGCTATAA 1254

RESULT 3

ADM48684

ID ADM48684 standard; DNA; 1254 BP.

XX AC ADM48684;

XX DT 03-JUN-2004 (first entry)

XX DE Human wild type rGAL-12 DNA.

XX KW Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;

XX KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;

XX KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;

XX KW gene; ds.

XX OS Homo sapiens.

XX FH Key

XX CDS Location/Qualifiers

FT 1..1254

FT /*tag= a

FT /product= "GAL-12 protein"

XX PN US2004023281-A1.

XX PD 05-FEB-2004.

XX PF 23-JUN-2003; 2003US-00602220.

XX PR 26-FEB-1988; 88US-00160766.

XX PR 26-FEB-1988; 88US-00160771.

XX PR 17-FEB-1989; 89US-00310881.

XX PR 22-OCT-1990; 90US-00600244.

XX PR 31-JUL-1992; 92US-00923692.

XX PR 30-DEC-1992; 92US-00997733.

XX PR 29-DEC-1993; 93US-00176414.

XX PR 19-JAN-1994; 94US-00184237.

XX PR 14-OCT-1994; 94US-00324003.

XX PR 21-MAY-1999; 99US-00316572.

XX PR 26-JUL-2000; 2000US-00626127.

XX PR 13-NOV-2001; 2001US-00993059.

XX (TUPR/) TURPEN T H.

XX (KUMA/) KUMAGAI M H.

XX (POGU/) POGUE G P.

XX (ERWL/) ERWIN R L.

XX (GRIL/) GRILL L K.

XX Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
XX WPI; 2004-142650/14.
XX P-PSDB; ADM48685.
XX New alpha-galactosidase polypeptides, useful in producing recombinant
PT lysosomal enzymes for the treatment of lysosomal storage diseases, such
PT as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
PT disease.
XX Disclosure; SEQ ID NO 15; 72pp; English.
XX The present invention relates to novel galactosidase (Gal) proteins such
CC as rGAL-12R, rGAL-25 or rGAL-25R. The methods and compositions
CC of the present invention are useful for producing recombinant lysosomal
CC enzymes for enzyme replacement therapy for treating human and animal
CC lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
CC disease, Fabry disease and Tay-Sachs disease. The present sequence is
CC human wild type rGAL-12 DNA used in the exemplification of the invention.
XX Sequence 1254 BP; 316 A; 289 C; 328 G; 321 T; 0 U; 0 Other;
Query Match 100.0%; Score 1254; DB 12; Length 1254;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGCCTTCCTTCGGCC 60
DB 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGCCTTCCTTCGGCC 60
QY 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACATGATGGCAAGGAGCGCT 120
DB 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACATGATGGCAAGGAGCGCT 120
QY 121 ACCATGCGTGGCTGGCACTGGGAGCGCTTCATGTGCAACTTGCCTGCCAGGAGGCCA 180
DB 121 ACCATGCGTGGCTGGCACTGGGAGCGCTTCATGTGCAACTTGCCTGCCAGGAGGCCA 180
QY 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
DB 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
QY 241 TGAAGGATGCGGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
DB 241 TGAAGGATGCGGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
QY 301 GATTTCAGAGGCGAGACTTCAGGCGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
DB 301 GATTTCAGAGGCGAGACTTCAGGCGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATATGTTTCACAGCAAGGAGCTGAAGCTAGGATTTATCAGATGTTGGAATAAA 420
DB 361 GCTAATATGTTTCACAGCAAGGAGCTGAAGCTAGGATTTATCAGATGTTGGAATAAA 420
QY 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATACCTAGACATGATGCCAGACTTTGCT 480
DB 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATACCTAGACATGATGCCAGACTTTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAAATTGATGTTGTTTACTGTGACAGTTTGGAAAAATTG 540
DB 481 GACTGGGAGTAGATCTGCTAAAATTGATGTTGTTTACTGTGACAGTTTGGAAAAATTG 540
QY 541 GCAGATGGTTTATAGCAGCATGTCTCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTC 600
DB 541 GCAGATGGTTTATAGCAGCATGTCTCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTC 600
QY 601 TCCCTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 660
DB 601 TCCCTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATGATGATTCCTGGAAAAAGTATAAAG 720
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATGATGATTCCTGGAAAAAGTATAAAG 720

DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATGATGATTCCTGGAAAAAGTATAAAG 720
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTGATGCTTCTGACACAGG 780
DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTGATGCTTCTGACACAGG 780
QY 781 GGTTCGAATGACCCAGATATGTTAGTGAATGGCACTTTGGGCTCAGCTGGAATCAGCAA 840
DB 781 GGTTCGAATGACCCAGATATGTTAGTGAATGGCACTTTGGGCTCAGCTGGAATCAGCAA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
QY 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCCAGTAAGAGACGTAATTCGCATCAAT 960
DB 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCCAGTAAGAGACGTAATTCGCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 1021 GAAAGACCTCTCTCAGGCTTACCTGGGCTGTAGCTATGATTAACCGGAGGAGATGCT 1080
DB 1021 GAAAGACCTCTCTCAGGCTTACCTGGGCTGTAGCTATGATTAACCGGAGGAGATGCT 1080
QY 1081 GGACCTCGCTCTTATACCATCCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAATCCT 1140
DB 1081 GGACCTCGCTCTTATACCATCCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAATCCT 1140
QY 1141 GCCTGCTTCATCACAGCTCCTCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
DB 1141 GCCTGCTTCATCACAGCTCCTCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 1201 TCAAGTTAAGAGTCACATAAATCCACAGGCACTGTTTTCCTTCAGCTATAA 1254
DB 1201 TCAAGTTAAGAGTCACATAAATCCACAGGCACTGTTTTCCTTCAGCTATAA 1254
RESULT 4
ADU66919 standard; DNA; 1254 BP.
XX AC ADU66919;
XX AC ADU66919;
XX 10-FEB-2005 (first entry)
XX Human alpha-galactosidase protein encoding DNA #7.
XX Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase;
KW enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;
KW Niemann-Pick disease; Fabry's disease; Tay-Sachs disease; Hurler's disease;
KW Hurler's syndrome; Hurler-Scheie syndrome; nephrotropic; human; gene; ds.
XX Homo sapiens.
XX Key Location/Qualifiers
PH 1. 1254
FT /*tag= a
FT /product= "Human alpha-galactosidase protein"
XX US2004234516-A1.
XX 25-NOV-2004.
XX 21-MAY-2004; 2004US-00851388.
XX 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-0093059.
PR 20-MAR-2002; 2002US-00103327.
XX (LARG-) LARGE SCALE BIOLOGY CORP.
XX PA
XX

XX (LARG-) LARGE SCALE BIOLOGY CORP.
XX PA
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPI: 2005-404004/41.
DR P-PDB; AEA27449.
XX
PT New isolated polypeptides useful for producing lysosomal enzymes in
PT plants to be utilized in enzyme replacement therapy or for the
PT therapeutic treatment of human or animal lysosomal storage diseases, e.g.
PT Gaucher's disease.
XX
PS Disclosure; SEQ ID NO 15; 88pp; English.
XX
CC The present invention relates to the production of human and animal
CC lysosomal enzymes in plants by a transient plant expression system. The
CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
CC galactosidase (Gal) enzymes having a post-translational modification
CC provided by the plant expression system. The invention is useful in
CC enzyme replacement therapy for treating lysosomal storage diseases such
CC as Gaucher's disease, Niemann-pick disease, Fabry's disease, Tay-Sachs
CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
CC also useful in researches for developing new approaches to medical
CC treatment of lysosomal storage diseases and in industrial processes
CC involving enzymatic substrate hydrolysis. The present sequence is the
CC human alpha-galactosidase DNA.
XX
SQ Sequence 1254 BP; 316 A; 289 C; 328 G; 321 T; 0 U; 0 Other;

Query Match 100.0%; Score 1254; DB 14; Length 1254;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 60
QY 61 CTCGTTTCTTGGACATCCCTGGGCTTAGAGCACTGGAACAATGGATGGCAAGGACGCT 120
DB 61 CTCGTTTCTTGGACATCCCTGGGCTTAGAGCACTGGAACAATGGATGGCAAGGACGCT 120
QY 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATCTGCAACCTTCAGCTGCCAGGAAGGCCA 180
DB 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATCTGCAACCTTCAGCTGCCAGGAAGGCCA 180
QY 181 GATTCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
DB 181 GATTCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 241 TGAAGGATGACGGTTATGATGACCTCTGCAATGATGACCTGTTGGATGGCTCCCAAGA 300
DB 241 TGAAGGATGACGGTTATGATGACCTCTGCAATGATGACCTGTTGGATGGCTCCCAAGA 300
QY 301 GATTGAGAGGACACTTCAGGAGACCCCTCAGCGCTTTCTCATGGATTCGCGAGCTA 360
DB 301 GATTGAGAGGACACTTCAGGAGACCCCTCAGCGCTTTCTCATGGATTCGCGAGCTA 360
QY 361 GCTAATTTATCTTCACAGCAAGGACTGAAGCTAGGAGATTTATGACATGTTGGAATAAA 420
DB 361 GCTAATTTATCTTCACAGCAAGGACTGAAGCTAGGAGATTTATGACATGTTGGAATAAA 420
QY 421 ACCTGCGCAGGCTTCCTGGAGTTTGGATACATGACATGATGATGATGATGATGATGAT 480
DB 421 ACCTGCGCAGGCTTCCTGGAGTTTGGATACATGACATGATGATGATGATGATGATGAT 480
QY 481 GACTGGGAGTACATCTGCTAATATTTGATGTTTGTGTTGTTGTTGTTGTTGTTGTTGTT 540
DB 481 GACTGGGAGTACATCTGCTAATATTTGATGTTTGTGTTGTTGTTGTTGTTGTTGTTGTT 540
QY 541 GCAGATGTTTATAGCACATGCTTCCTGGCCCTGAAATAGGAGCTGGCAGAAGCATTTGTAC 600
DB 541 GCAGATGTTTATAGCACATGCTTCCTGGCCCTGAAATAGGAGCTGGCAGAAGCATTTGTAC 600

QY 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGGCCAAATTTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGGCCAAATTTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCAGTGGCGAAATTTTGTGTCATTTGTCGATGATGATGATGATGATGAT 720
DB 661 CAGTACTGCAATCAGTGGCGAAATTTTGTGTCATTTGTCGATGATGATGATGATGATGAT 720
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGCTGCTGACACAGGG 780
DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGCTGCTGACACAGGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTATGTTGGCACTTTGGCCCTCAGCTGGAAATCAGCA 840
DB 781 GGTGGAATGACCCAGATATGTTAGTATGTTGGCACTTTGGCCCTCAGCTGGAAATCAGCA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATGTCATGTCATGACCTC 900
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATGTCATGTCATGACCTC 900
QY 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
DB 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACCTTTGAAGTGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACCTTTGAAGTGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTTAGCTATGATATAACCGGAGGAGATTGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTTAGCTATGATATAACCGGAGGAGATTGGT 1080
QY 1081 GGAACCTGCTCTTATACCATCGCAGTTGCTTCCTGGGTAAAGGAGTGGCTCTGTAATCCT 1140
DB 1081 GGAACCTGCTCTTATACCATCGCAGTTGCTTCCTGGGTAAAGGAGTGGCTCTGTAATCCT 1140
QY 1141 GCCTGCTTCATACACAGCTCCTCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
DB 1141 GCCTGCTTCATACACAGCTCCTCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTTGCTTCAGCTATAA 1254
DB 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTTGCTTCAGCTATAA 1254

RESULT 6
AAD45223
ID AAD45223 standard; DNA; 1266 BP.
XX
AC AAD45223;
XX 27-DEC-2002 (first entry)
DT
XX Human rGAL-8 DNA.
DE
XX Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
XX therapeutic; rGAL-8; gene; ds.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..1266
XX FT /*tag= a
XX FT /product= "Human rGAL-8 protein"
XX
XX US2002088024-A1.
XX
XX 04-JUL-2002.
XX
XX 13-NOV-2001; 2001US-00993059.
XX
XX 26-JUL-2000; 2000US-00626127.
XX
PR

XX (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2002-681656/73.
DR P-PSDB; AAE28210.
XX Novel human alpha-galactosidase polypeptide useful for treating lysosomal
PT storage diseases.
XX
XX Claim 1; Page 42-44; 88pp; English.
XX The invention relates to human alpha-galactosidase truncated at the
CC carboxy terminus and the production of enzymatically active recombinant
CC human and animal lysosomal enzymes. The invention is useful for producing
CC lysosomal enzymes for treating lysosomal storage diseases, producing
CC altered or mutated proteins, enzymatically active or otherwise, to serve
CC as precursors or substrates for further in vivo or in vitro processing to
CC a specialised industrial form for research or therapeutic uses, to
CC produce more effective therapeutic enzyme, for producing antibodies
CC against lysosomal enzymes for medical diagnostic use, and in any
CC commercial process that involves substrate hydrolysis. The present
CC sequence is human rGAL-8 DNA
XX
SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;

Query Match 99.98; Score 1252.4; DB 6; Length 1266;
Best Local Similarity 99.98; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCTTGGCTTGGCTTCTTCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCTTGGCTTGGCTTCTTCTGGCC 60
QY 61 CTGCTTCTTGGGACATCTCTGGGCTGAGAGCACTGGCAATGGATGGCAAGAGCGCT 120
DB 61 CTGCTTCTTGGGACATCTCTGGGCTGAGAGCACTGGCAATGGATGGCAAGAGCGCT 120
QY 121 ACCATGGGCTGCTGCACTGGGAGCGCTTCATCTGCAACCTTCACTGCCAGGAGAGCCA 180
DB 121 ACCATGGGCTGCTGCACTGGGAGCGCTTCATCTGCAACCTTCACTGCCAGGAGAGCCA 180
QY 181 GATTCTGTCATCAGTGAGAGAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
DB 181 GATTCTGTCATCAGTGAGAGAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 241 TGAAGAGATGAGGTTATGATGATACCTCTGCAATTGATGACTGTTGGATGGCTCCCCAAAGA 300
DB 241 TGAAGAGATGAGGTTATGATGATACCTCTGCAATTGATGACTGTTGGATGGCTCCCCAAAGA 300
QY 301 GATTGAGAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTCTTCATGGGATTCGCCAGCTA 360
DB 301 GATTGAGAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTCTTCATGGGATTCGCCAGCTA 360
QY 361 GCTAAATTATGTTTACAGCAAGAGACTGAAGCTAGGAGATTTATGAGATGTTGAAATATAA 420
DB 361 GCTAAATTATGTTTACAGCAAGAGACTGAAGCTAGGAGATTTATGAGATGTTGAAATATAA 420
QY 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACCTAGCAGATTTGATGCCAGACTTTGGCT 480
DB 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACCTAGCAGATTTGATGCCAGACTTTGGCT 480
QY 481 GACTGGGGAGTAGACTGCTGCTAAATTTGATGTTTACTGTGACAGATTTGGAAATTTG 540
DB 481 GACTGGGGAGTAGACTGCTGCTAAATTTGATGTTTACTGTGACAGATTTGGAAATTTG 540
QY 541 GCAGATGGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCACTTGTGTAC 600
DB 541 GCAGATGGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCACTTGTGTAC 600

QY 601 TCCTGTGAGTGGCCTCTTTATATGTGGCCTTTCAAAGCCCAATTATATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCCTCTTTATATGTGGCCTTTCAAAGCCCAATTATATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTGATGATTCCTGGAAAAGTATAAAG 720
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTGATGATTCCTGGAAAAGTATAAAG 720
QY 721 AGTATCTTGAGACTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGTGACCAAGG 780
DB 721 AGTATCTTGAGACTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGTGACCAAGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCTCAGCTGGAATCAGCA 840
DB 781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCTCAGCTGGAATCAGCA 840
QY 841 GTAACCTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
DB 841 GTAACCTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACCACTTAGACAGGAGACAACTTTGAAGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGTTACCACTTAGACAGGAGACAACTTTGAAGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGATGATATAAACCGGACGAGATGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGATGATATAAACCGGACGAGATGGT 1080
QY 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGTAAAGAGTGGCTGTAAATCCT 1140
DB 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGTAAAGAGTGGCTGTAAATCCT 1140
QY 1141 GCCTGCTTATCACAGCTCTCTCCCTGTGAAAAGAGTAGGTTCTATGATGAGCT 1200
DB 1141 GCCTGCTTATCACAGCTCTCTCCCTGTGAAAAGAGTAGGTTCTATGATGAGCT 1200
QY 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254
DB 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254

RESULT 7
ADD84750
ID ADD84750 standard; DNA; 1266 BP.
XX
AC ADD84750;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human alpha-galactosidase rGAL-8 DNA.
XX
KW Human; alpha-galactosidase; rGAL-4; gene; ds; lysosomal enzyme;
KW enzyme replacement therapy; lysosomal disease.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 1..1266
FT /*tag= a
FT /product= "Human rGAL-8"
XX
PN US2003106095-A1.
XX
PD 05-JUN-2003.
XX
XX 20-MAR-2002; 2002US-00103327.
XX
XX 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.

```
XX (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
XX
PI Garger SJ, Turpen TH, Kumagai MH;
XX
XX WPI; 2003-801257/75.
DR P-PSDB; ADD84751.
XX
XX New polynucleotide for producing active recombinant human and animal
PT lysosomal enzymes in a plant expression system that can be used in enzyme
PT replacement therapy.
XX
XX Claim 1; SEQ ID NO 11; 77pp; English.
XX
XX The invention relates to human alpha-galactosidase derivatives and the
CC nucleic acids encoding them. The polypeptides are used in a method for
CC producing active recombinant human and animal lysosomal enzymes in a
CC plant expression system. The enzymes can be used in enzyme replacement
CC therapy for the therapeutic treatment of human and animal lysosomal
CC diseases. This sequence represents DNA encoding a human alpha-
CC galactosidase derivative polypeptide of the invention.
XX
XX Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;
SQ
Query Match 99.9%; Score 1252.4; DB 10; Length 1266;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGNACCAGACTACATCTGGGCTGGCGCTGGCGCTTCGCTTCCTGGCC 60
DB 1 ATGCAGCTGAGNACCAGACTACATCTGGGCTGGCGCTGGCGCTTCGCTTCCTGGCC 60
QY 61 CTGCTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGAGCGCT 120
DB 61 CTGCTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGAGCGCT 120
QY 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
DB 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
QY 181 GATTCTGCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAGGC 240
DB 181 GATTCTGCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAGGC 240
QY 241 TGAAGGATGCAGGTTATGATGATACCTCTGATTTGATGATGCTTGGATGGCTCCCAAGA 300
DB 241 TGAAGGATGCAGGTTATGATGATACCTCTGATTTGATGATGCTTGGATGGCTCCCAAGA 300
QY 301 GATTGAGAGGCGAGCTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
DB 301 GATTGAGAGGCGAGCTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATATGTTACAGCAAGAGCTGAGCTAGGAGTTATCGAGATGTTGGAATAAA 420
DB 361 GCTAATATGTTACAGCAAGAGCTGAGCTAGGAGTTATCGAGATGTTGGAATAAA 420
QY 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGGATCTACGACATGATGCCAGAGCTTTGCT 480
DB 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGGATCTACGACATGATGCCAGAGCTTTGCT 480
QY 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTG 540
DB 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTG 540
QY 541 GCAGATGGTTATAAGACATGCTCTGGCCCTGAATAGGACTGCGAAGCATTTGTGTAC 600
DB 541 GCAGATGGTTATAAGACATGCTCTGGCCCTGAATAGGACTGCGAAGCATTTGTGTAC 600
QY 601 TCTGTGAGTGGCTCTTTATGATGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
DB 601 TCTGTGAGTGGCTCTTTATGATGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
```

```
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTGATGATTCCTGGAAGATATAAG 720
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTGATGATTCCTGGAAGATATAAG 720
QY 721 AGTATCTTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTGCTGGACAGGG 780
DB 721 AGTATCTTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTGCTGGACAGGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTATGTCACACTTTGGCCTCAGCTCGGAATCAGCA 840
DB 781 GGTGGAATGACCCAGATATGTTAGTATGTCACACTTTGGCCTCAGCTCGGAATCAGCA 840
QY 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCATTAATGACCTC 900
DB 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCATTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGCCATCAAT 960
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGAGCAACTTTCGAGGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGAGCAACTTTCGAGGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAAACCGGAGGAGATTGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAAACCGGAGGAGATTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCGAGTTGCTCCCTGGGTAAGGAGTGGCTGTAAATCCT 1140
DB 1081 GGACCTCGCTCTTATACCATCGAGTTGCTCCCTGGGTAAGGAGTGGCTGTAAATCCT 1140
QY 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATCGACT 1200
DB 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATCGACT 1200
QY 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTGTCTTCAGCTATAA 1254
DB 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTGTCTTCAGCTATAA 1254
RESULT 8
ADJ88276
ID ADJ88276 standard; DNA; 1266 BP.
XX
AC ADJ88276;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human WT rGAL-8 (galactosidase) DNA.
XX
XX Galactosidase; GAL; gene therapy; lysosomal storage disease;
KW Fabry's disease; Gaucher's disease; human; gene; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
CDS 1..1266
FT /*tag= a
FT /product= "Human WT rGAL-8 protein"
XX
PN US2004016021-A1.
XX
XX 22-JAN-2004.
XX
XX 23-JUN-2003; 2003US-00602219.
XX
XX 26-FEB-1988; 88US-00160766.
PR 26-FEB-1988; 88US-00160771.
PR 15-JUL-1988; 88US-00219279.
PR 17-FEB-1989; 89US-00310881.
PR 05-MAY-1989; 89US-00347637.
```

PR 08-JUN-1989; 89US-00363138.
 PR 22-OCT-1990; 90US-00600244.
 PR 16-JAN-1991; 91US-00641617.
 PR 26-JUL-1991; 91US-00737899.
 PR 01-AUG-1991; 91US-00739143.
 PR 31-JUL-1992; 92US-00923692.
 PR 30-DEC-1992; 92US-00997733.
 PR 29-DEC-1993; 93US-00176414.
 PR 19-JAN-1994; 94US-00184237.
 PR 14-OCT-1994; 94US-00324003.
 PR 21-MAY-1999; 99US-00316572.
 PR 26-JUL-2000; 2000US-00626127.
 PR 13-NOV-2001; 2001US-00993059.
 XX (TURP/) TURPEN T H.
 PA (POGU/) POGUE G P.
 PA (ERWI/) ERWIN R L.
 PA (GRIL/) GRILL L K.
 XX
 PI Turpen TH, Pogue GP, Erwin RL, Grill LK;
 XX
 DR WPI; 2004-108227/11.
 DR P-PSDB; ADJ88277.
 XX
 PT New lysosomal enzymes, useful in treating human and animal lysosomal
 PT storage diseases, e.g. Fabry's disease and Gaucher's diseases.
 XX
 PS Claim 1; SEQ ID NO 11; 71pp; English.
 XX
 CC The invention relates to nucleotide encoding galactosidase (GAL). The
 CC invention is useful in gene therapy. The polynucleotides and polypeptides
 CC are useful in treating human and animal lysosomal storage diseases, e.g.
 CC Fabry's disease and Gaucher's diseases. The present sequence is human GAL
 CC DNA.
 XX
 SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;
 Query Match 99.9%; Score 1252.4; DB 12; Length 1266;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTTCCTGGCC 60
 DB 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTTCCTGGCC 60
 QY 61 CTCGTTTCTCGGACATCCCTCGGGCTAGAGCACTGGCAATGGATTGGCAAGACGCT 120
 DB 61 CTCGTTTCTCGGACATCCCTCGGGCTAGAGCACTGGCAATGGATTGGCAAGACGCT 120
 QY 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
 DB 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
 QY 181 GATTCTGTCATCAGTGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
 DB 181 GATTCTGTCATCAGTGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
 QY 241 TGAAGGATGACAGGTATAGTACCTCTGCAATGATGACATGTTGGATGGCTCCCAAGA 300
 DB 241 TGAAGGATGACAGGTATAGTACCTCTGCAATGATGACATGTTGGATGGCTCCCAAGA 300
 QY 301 GATTTCAGAGGACAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGATTCGCCAGCTA 360
 DB 301 GATTTCAGAGGACAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGATTCGCCAGCTA 360
 QY 361 GCTAAATATGTTACAGCAAGAGCTGGAAGCTAGGAGATTATGAGATGTTGGAATATAA 420
 DB 361 GCTAAATATGTTACAGCAAGAGCTGGAAGCTAGGAGATTATGAGATGTTGGAATATAA 420
 QY 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACCTAGCAGATTGATGCCAGAGCTTTGCT 480
 DB 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACCTAGCAGATTGATGCCAGAGCTTTGCT 480

QY 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGACAGTTGGAAAAATTTG 540
 DB 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGACAGTTGGAAAAATTTG 540
 QY 541 GCAGATGTTTATAAGACACATGCTCTGGCCCTGAATAGGACTGCGCAGAACGATTGTGTAC 600
 DB 541 GCAGATGTTTATAAGACACATGCTCTGGCCCTGAATAGGACTGCGCAGAACGATTGTGTAC 600
 QY 601 TCCTGTAGTGGCTCTTTTATATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 660
 DB 601 TCCTGTAGTGGCTCTTTTATATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 660
 QY 661 CAGTACTGCAATCACTGGCGGAAATTTTGTGACATTTGATGATGATTTCTCGGAAAGATATAAG 720
 DB 661 CAGTACTGCAATCACTGGCGGAAATTTTGTGACATTTGATGATGATTTCTCGGAAAGATATAAG 720
 QY 721 AGTATCTTGACCTGGACATCTTTTAACAGGAGAGAAATCTGTGATGTTGCTGGACAGGG 780
 DB 721 AGTATCTTGACCTGGACATCTTTTAACAGGAGAGAAATCTGTGATGTTGCTGGACAGGG 780
 QY 781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
 DB 781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
 QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTTATTCATGCTTAATGACCTC 900
 DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTTATTCATGCTTAATGACCTC 900
 QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
 DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
 QY 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAAGCTTTAGACAGGAGAGCAACTTTGAAGTGTGG 1020
 DB 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAAGCTTTAGACAGGAGAGCAACTTTGAAGTGTGG 1020
 QY 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATATAACCGGAGGAGATTGGT 1080
 DB 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATATAACCGGAGGAGATTGGT 1080
 QY 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCTGTAAATCCT 1140
 DB 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCTGTAAATCCT 1140
 QY 1141 GCCTGCTTCATCACACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
 DB 1141 GCCTGCTTCATCACACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
 QY 1201 TCAAGGTTAAGAGTCAATAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254
 DB 1201 TCAAGGTTAAGAGTCAATAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254

RESULT 9

ADM48680

ID ADM48680 standard; DNA; 1266 BP.

XX

AC ADM48680;

XX

DT 03-JUN-2004 (first entry)

XX

DE Human wild type rGAL-8 DNA.

XX

KW Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;

KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;

KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;

XX gene; ds.

OS Homo sapiens.

XX

PH Key Location/Qualifiers

FT CDS 1..1266

/*tag= a

```
FT /product= "GAL-8 protein"
XX US2004023281-A1.
XX 05-FEB-2004.
XX 23-JUN-2003; 2003US-00602220.
XX 26-FEB-1988; 88US-00160766.
XX 26-FEB-1988; 88US-00160771.
XX 17-FEB-1989; 89US-00310881.
XX 22-OCT-1990; 90US-00600244.
XX 31-JUL-1992; 92US-00923692.
XX 30-DEC-1992; 92US-00929773.
XX 29-DEC-1993; 93US-00176414.
XX 19-JAN-1994; 94US-00184237.
XX 14-OCT-1994; 94US-00324003.
XX 21-MAY-1999; 99US-00316572.
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.
XX (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX (POGU/) POGUE G P.
XX (ERWL/) ERWIN R L.
XX (GRIL/) GRILL L K.
XX Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
XX WPI; 2004-142650/14.
XX F-PSDB; ADM48681.
XX New alpha-galactosidase polypeptides, useful in producing recombinant
XX lysosomal enzymes for the treatment of lysosomal storage diseases, such
XX as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
XX disease.
XX Disclosure; SEQ ID NO 11; 72pp; English.
XX The present invention relates to novel galactosidase (Gal) proteins such
XX as rGAL-12R, rGAL-25 or rGAL-25R. The methods and compositions
XX of the present invention are useful for producing recombinant lysosomal
XX enzymes for enzyme replacement therapy for treating human and animal
XX lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
XX disease, Fabry disease and Tay-Sachs disease. The present sequence is
XX human wild type rGAL-8 DNA used in the exemplification of the invention.
XX Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;
XX
XX Query Match 99.9%; Score 1252.4; DB 12; Length 1266;
XX Best Local Similarity 99.9%; Pred. No. 0;
XX Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGCGCTTCGCTTCCTGGCC 60
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGCGCTTCGCTTCCTGGCC 60
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 61 CTGCTTTCTTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGGCGCTT 120
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 61 CTGCTTTCTTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGGCGCTT 120
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 181 GATTCTCGCATCAGTAGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 181 GATTCTCGCATCAGTAGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 241 TGGAAAGGATCAGGTTATGAGTACCTCTGCAATTGATGACTGTGTGGATGGCTCCCAAAGA 300
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 241 TGGAAAGGATCAGGTTATGAGTACCTCTGCAATTGATGACTGTGTGGATGGCTCCCAAAGA 300
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
QY 301 GATTTCAGAGGCAGACTTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 301 GATTTCAGAGGCAGACTTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 361 GCTAATTATGTTTCACAGCAAGGACTGAAGCTAGGATTTATGCAGATGTTGGAAATAA 420
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 361 GCTAATTATGTTTCACAGCAAGGACTGAAGCTAGGATTTATGCAGATGTTGGAAATAA 420
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACTACGACATTTGATGATCCCGACAGCTTTGCT 480
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACTACGACATTTGATGATCCCGACAGCTTTGCT 480
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 481 GACTGGGAGTAGACTCTGCTAAAATTTGATGTTGTTACTGTGACAGTTTGGAAAATTG 540
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 481 GACTGGGAGTAGACTCTGCTAAAATTTGATGTTGTTACTGTGACAGTTTGGAAAATTG 540
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 541 GCAGATGGTTATAAGCACAATGCTTGGCCCTGGAATAGACTGGCAGAGACATTTGTGATC 600
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 541 GCAGATGGTTATAAGCACAATGCTTGGCCCTGGAATAGACTGGCAGAGACATTTGTGATC 600
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 601 TCCTGTGAGTGGCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 601 TCCTGTGAGTGGCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 661 CAGTACTGCAATCAGCTGGCGAAATTTTGTGCAATTTGATGATTTCTGGAAAAGTATAAAG 720
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 661 CAGTACTGCAATCAGCTGGCGAAATTTTGTGCAATTTGATGATTTCTGGAAAAGTATAAAG 720
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGGAGAAATTTGTGATGTTGCTGACACAGG 780
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGGAGAAATTTGTGATGTTGCTGACACAGG 780
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 781 GGTGGAATGACCCAGATATTTAGTATTTGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 781 GGTGGAATGACCCAGATATTTAGTATTTGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTTAAAGACCTC 900
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTTAAAGACCTC 900
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATGGCCATCAAT 960
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATGGCCATCAAT 960
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1021 GAACGACCTCTCTCAGGCTTTAGCTGGGCTGTAGCTATGATATAACCGGCAGGAGATTGGT 1080
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1021 GAACGACCTCTCTCAGGCTTTAGCTGGGCTGTAGCTATGATATAACCGGCAGGAGATTGGT 1080
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1081 GGAACCTCGCTCTTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTGTAACTCT 1140
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1081 GGAACCTCGCTCTTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTGTAACTCT 1140
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1141 GCCTGCTTCATCAGACAGCTCTCCCTGTGAAAAGGAAGCTAGGGTCTCTATGATGAGCT 1200
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1141 GCCTGCTTCATCAGACAGCTCTCCCTGTGAAAAGGAAGCTAGGGTCTCTATGATGAGCT 1200
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1201 TCAAGGTTAAGAAAGTACATAAAATCCCAAGGCACTGTGTTTGTCTCAGCTATAA 1254
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 1201 TCAAGGTTAAGAAAGTACATAAAATCCCAAGGCACTGTGTTTGTCTCAGCTATAA 1254
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
XX RESULT 10
XX ADU66915
XX ID ADU66915 standard; DNA; 1266 BP.
XX AC ADU66915;
XX DT 10-FEB-2005 (first entry)
XX XX
```



```
XX 11-AUG-2005 (first entry)
XX Human alpha-galactosidase DNA, rGAL-8, SEQ ID NO: 11.
XX
XX Gauchers disease; metabolic; neurological disease; niemann pick disease;
XX Genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
XX antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase;
XX gene; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..1266
XX /tag= a
XX /product= "Human alpha-galactosidase protein"
XX
XX US2005125859-A1.
XX
XX 09-JUN-2005.
XX
XX 08-NOV-2004; 2004US-00984389.
XX
XX 26-JUL-2000; 2000US-00626127.
XX
XX 13-NOV-2001; 2001US-00993059.
XX
XX 20-MAR-2002; 2002US-00103327.
XX
XX (LARG-) LARGE SCALE BIOLOGY CORP.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
XX
XX WPI; 2005-404004/41.
XX
XX P-PSDB; AEA27445.
XX
XX New isolated polypeptides useful for producing lysosomal enzymes in
XX plants to be utilized in enzyme replacement therapy or for the
XX therapeutic treatment of human or animal lysosomal storage diseases, e.g.
XX Gaucher's disease.
XX
XX Disclosure; SEQ ID NO 11; 88pp; English.
XX
XX The present invention relates to the production of human and animal
XX lysosomal enzymes in plants by a transient plant expression system. The
XX invention relates to glucocerebrosidase (GCB, GCR) and alpha-
XX galactosidase (Gal) enzymes having a post-translational modification
XX provided by the plant expression system. The invention is useful in
XX enzyme replacement therapy for treating lysosomal storage diseases such
XX as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
XX disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
XX also useful in researches for developing new approaches to medical
XX treatment of lysosomal storage diseases and in industrial processes
XX involving enzymatic substrate hydrolysis. The present sequence is the
XX human alpha-galactosidase DNA.
XX
XX Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;
XX
XX
XX Query Match 99.9%; Score 1252.4; DB 14; Length 1266;
XX Best Local Similarity 99.9%; Pred. No. 0;
XX Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 ATCGAGCTGAGGACCCAGAACTACATCTGGGCTGGCGCTGGCGCTTCGCTTCCTGGCC 60
XX |
XX DB 1 ATCGAGCTGAGGACCCAGAACTACATCTGGGCTGGCGCTGGCGCTTCGCTTCCTGGCC 60
XX |
XX QY 61 CTCGTTTCCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGAATGGCAAGACGCTT 120
XX |
XX DB 61 CTCGTTTCCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGAATGGCAAGACGCTT 120
XX |
XX QY 121 ACCATGGGCTGGCTGCACTGGGACGCTTCATGTGCAACCTTGATGCCAGGAGGCCA 180
XX |
XX DB 121 ACCATGGGCTGGCTGCACTGGGACGCTTCATGTGCAACCTTGATGCCAGGAGGCCA 180
XX |
XX QY 181 GATTCTGCATCAGTGAAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
XX |
```

RESULT 12


```
XX 06-MAY-2004 (first entry)
XX DE Human rGAL- 4 (galactosidase) DNA.
XX
XX Galactosidase; GAL; gene therapy; lysosomal storage disease;
XX Fabry's disease; Gaucher's disease; human; gene; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..1278
XX /*cag= a
XX /product= "Human WT rGAL-4 protein"
XX /partial
XX /note= "No start codon"
XX
XX US2004016021-A1.
XX
XX 22-JAN-2004.
XX
XX 23-JUN-2003; 2003US-00602219.
XX
XX 26-FEB-1988; 88US-00160766.
XX 26-FEB-1988; 88US-00160771.
XX 15-JUL-1988; 88US-00219279.
XX 17-FEB-1989; 89US-00310881.
XX 05-MAY-1989; 89US-00347637.
XX 08-JUN-1989; 89US-00363138.
XX 22-OCT-1990; 90US-00600244.
XX 16-JAN-1991; 91US-00641617.
XX 26-JUL-1991; 91US-00737899.
XX 01-AUG-1991; 91US-00739143.
XX 31-JUL-1992; 92US-00923692.
XX 30-DEC-1992; 92US-00997733.
XX 29-DEC-1993; 93US-00176414.
XX 19-JAN-1994; 94US-00184237.
XX 14-OCT-1994; 94US-00324003.
XX 21-MAY-1999; 99US-00316572.
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.
XX
XX (TURP/) TURPEN T H.
XX (FOGU/) FOGUE G P.
XX (ERWI/) ERWIN R L.
XX (GRIL/) GRILL L K.
XX
XX Turpen TH, Pogue GP, Erwin RL, Grill LK;
XX WPI; 2004-108227/11.
XX P-PSDB; ADJ88273.
XX
XX New lysosomal enzymes, useful in treating human and animal lysosomal
XX storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
XX Claim 1; SEQ ID NO 7; 71bp; English.
XX
XX The invention relates to nucleotide encoding galactosidase (GAL). The
XX invention is useful in gene therapy. The polynucleotides and polypeptides
XX are useful in treating human and animal lysosomal storage diseases, e.g.
XX Fabry's disease and Gaucher's diseases. The present sequence is human
XX DNA.
XX
XX Sequence 1278 BP; 325 A; 293 C; 333 G; 327 T; 0 U; 0 Other;
XX
XX Query Match 99.9%; Score 1252.4; DB 12; Length 1278;
XX Best Local Similarity 99.9%; Pred. No. 0;
XX Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 ATGCAGCTGGAGAACCCAGACTACATCTGGGCTGCGGCTTGGCTTCCTGCTTCCTGGCC 60
XX |
XX 1 ATGCAGCTGGAGAACCCAGACTACATCTGGGCTGCGGCTTGGCTTCCTGCTTCCTGGCC 60
XX |
XX GCCTGCTTCATCACAAGCTCCTCCCTGTGAAAGGAAGCTAGGGTCTTATGATGAGTACT 1200
```

Db 1141 GCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAGCTAGGGTTCTATGATGACT 1200
QY 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTAGAA 1254

RESULT 15

ADM48676
ID ADM48676 standard; DNA; 1278 BP.
XX AC ADM48676;
XX 03-JUN-2004 (first entry)
DT XX
DE Human wild type rGAL-4 DNA.
XX Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
gene; ds.
XX OS Homo sapiens.

PH Key Location/Qualifiers
FT CDS 4..1278
FT /*tag= a
FT /product= "GAL-4 protein"
FT /partial
FT /note= "No start codon"

US2004023281-A1.

XX 05-FEB-2004.
XX 23-JUN-2003; 2003US-00602220.
XX 26-FEB-1988; 88US-00160766.
XX 26-FEB-1988; 88US-00160771.
XX 17-FEB-1989; 89US-00310881.
XX 22-OCT-1990; 90US-00600244.
XX 31-JUL-1992; 92US-00923692.
XX 30-DEC-1992; 92US-00997733.
XX 29-DEC-1993; 93US-00176414.
XX 19-JAN-1994; 94US-00184237.
XX 14-OCT-1994; 94US-00324003.
XX 21-MAY-1999; 99US-00316572.
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.

XX (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX (POGU/) POGUE G P.
XX (ERWI/) ERWIN R L.
XX (GRIL/) GRILL L K.

Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;

XX WPI; 2004-142650/14.
XX P-PSDB; ADM48677.

XX New alpha-galactosidase polypeptides, useful in producing recombinant
PT lysosomal enzymes for the treatment of lysosomal storage diseases, such
as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
disease.

Disclosure; SEQ ID NO 7; 72pp; English.

XX The present invention relates to novel galactosidase (Gal) proteins such
as rGAL-12R, rGAL-12S or rGAL-25R. The methods and compositions
of the present invention are useful for producing recombinant lysosomal
enzymes for enzyme replacement therapy for treating human and animal

CC lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
CC disease, Fabry disease and Tay-Sachs disease. The present sequence is
CC human wild type rGAL-4 DNA used in the exemplification of the invention.

XX Sequence 1278 BP; 325 A; 293 C; 333 G; 327 T; 0 U; 0 Other;

QY Query Match 99.9%; Score 1252.4; DB 12; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTCCCTGGCC 60

Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTCCCTGGCC 60

QY 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGCAAAATGATTTGGCAAGGACGCT 120

Db 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGCAAAATGATTTGGCAAGGACGCT 120

QY 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGCACGGGAAGAGCCCA 180

Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGCACGGGAAGAGCCCA 180

QY 181 GATTCCTGCATCAGTGAGAACTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240

Db 181 GATTCCTGCATCAGTGAGAACTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240

QY 241 TGGAGGATGCAGGTTATGAGTACCTCTGCATGTAGTACTGTTGGATGGCTCCCCAAAGA 300

Db 241 TGGAGGATGCAGGTTATGAGTACCTCTGCATGTAGTACTGTTGGATGGCTCCCCAAAGA 300

QY 301 GATTCAGAGGCGAGACTTCAGGCAGACCCCTCAGCGCTTCTCATGGATTCGCCAGCTA 360

Db 301 GATTCAGAGGCGAGACTTCAGGCAGACCCCTCAGCGCTTCTCATGGATTCGCCAGCTA 360

QY 361 GCTAATATGTTTCACAGCAAGGACTGAAGCTAGGAGTTATCAGATGTTGGAAATAA 420

Db 361 GCTAATATGTTTCACAGCAAGGACTGAAGCTAGGAGTTATCAGATGTTGGAAATAA 420

QY 421 ACCTGCGCAGGCTTCCTGGGAGTTTGGATACCTACGACATTCATGCCACAGACTTTGCT 480

Db 421 ACCTGCGCAGGCTTCCTGGGAGTTTGGATACCTACGACATTCATGCCACAGACTTTGCT 480

QY 481 GACTGGGAGTAGACTCTGCTAAAATTTGATGGTGTACTGTGACAGTTTGGAAAAATTG 540

Db 481 GACTGGGAGTAGACTCTGCTAAAATTTGATGGTGTACTGTGACAGTTTGGAAAAATTG 540

QY 541 GCAGATGTTTAAAGCACATGCTTGGCCCTCAATAGGACTGGCAGAGCATTTGTGTAC 600

Db 541 GCAGATGTTTAAAGCACATGCTTGGCCCTCAATAGGACTGGCAGAGCATTTGTGTAC 600

QY 601 TCCTGTGAGTGGCTCTTTTATATGTCGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660

Db 601 TCCTGTGAGTGGCTCTTTTATATGTCGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660

QY 661 CAGTACTGCAATCACTGGCGAAAATTTTGTGACATTTGATGATTCCTGGAAAAAGTATAAG 720

Db 661 CAGTACTGCAATCACTGGCGAAAATTTTGTGACATTTGATGATTCCTGGAAAAAGTATAAG 720

QY 721 AGTATCTTGACATGGACATCTTTTAAACCAGGAGAAATTTGATGTTGTCGACCCAGGG 780

Db 721 AGTATCTTGACATGGACATCTTTTAAACCAGGAGAAATTTGATGTTGTCGACCCAGGG 780

QY 781 GGTGTGAATCACCAGATATGTTAGTATGTCGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840

Db 781 GGTGTGAATCACCAGATATGTTAGTATGTCGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840

QY 841 GTAACTCAGATGGCCCTCTCGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900

Db 841 GTAACTCAGATGGCCCTCTCGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900

QY 901 CGACATCAGCCCTCAGGCAAGCTCTCTTCAGGATAAGGACGTAATGGCCATCAAT 960

Db 901 CGACATCAGCCCTCAGGCAAGCTCTCTTCAGGATAAGGACGTAATGGCCATCAAT 960

```
Qy 961 CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
Db |||||
Qy 961 CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
Db |||||
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATAAACCGGACAGGAGATTGGT 1080
Db |||||
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATAAACCGGACAGGAGATTGGT 1080
Db |||||
Qy 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Db |||||
Qy 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Db |||||
Qy 1141 GCCTGTTCATCAACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Db |||||
Qy 1141 GCCTGTTCATCAACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Db |||||
Qy 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACCTGTTTGTCTTCAGCTATAA 1254
Db |||||
Qy 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACCTGTTTGTCTTCAGCTAGAA 1254
Db |||||
```

Search completed: December 31, 2005, 17:36:17
Job time : 848 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 31, 2005, 15:58:11 ; Search time 5614 Seconds
(without alignments)
10450.837 Million cell updates/sec

Title: US-10-602-220-15
Perfect score: 1254
Sequence: 1 atgcagctgaggaaccaga.....ctgtttgttcagctataa 1254

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1:	gb_est1:*
2:	gb_est2:*
3:	gb_est3:*
4:	gb_hic:*
5:	gb_est4:*
6:	gb_est5:*
7:	gb_est6:*
8:	gb_est7:*
9:	gb_gss1:*
10:	gb_gss2:*
11:	gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1252.4	99.9	1266	4	CR607242	full-leng
2	1252.4	99.9	1277	4	CR605654	full-leng
3	1248	99.5	1253	4	CR617861	full-leng
4	1160.4	92.5	1290	10	AY408540	Homo sapi
5	1140	90.9	1290	10	AY408541	Homo sapi
6	968.6	77.2	1086	1	AL554978	AY408541 Pan trogl
7	926.8	73.9	1014	5	BX354096	AL554978
8	922.8	73.6	999	1	AL552630	BX354096
9	915.8	73.0	1067	3	BM564282	AL552630
10	915	73.0	1026	1	AL577581	AGENCOURT
11	897.2	71.5	1376	4	AK054547	AL577581
12	897.2	71.5	2962	4	AK040081	Mus muscu
13	883.2	70.4	922	7	CO645623	Mus muscu
14	881.8	70.3	1133	7	CO645672	ILLUMIGEN
15	874.8	69.8	1071	1	AL575861	ILLUMIGEN
16	874.6	69.7	1005	3	BQ062192	AL575861
17	870	69.4	984	7	CO646251	AGENCOURT
18	862.4	68.8	920	7	CO645464	ILLUMIGEN
19	856.4	68.3	923	7	CT005156	ILLUMIGEN
20	826	65.9	1296	10	AY408542	CT005156
21	814.2	64.9	976	5	BQ956043	Mus muscu
22	807.8	64.4	937	5	BQ934640	AGENCOURT

23	797	63.6	922	5	BUS40848	AGENCOURT
24	794.8	63.4	898	5	BUI191867	AGENCOURT
25	783.8	62.5	888	5	BUI154569	AGENCOURT
26	782	62.4	852	6	CA454083	AGENCOURT
27	768	61.2	801	5	BUS96617	AGENCOURT
28	765.4	61.0	927	2	BE622583	AGENCOURT
29	765.2	61.0	813	5	BX344841	AGENCOURT
30	761.4	60.7	871	6	CA454143	AGENCOURT
31	757.6	60.4	1025	3	BM450649	AGENCOURT
32	752.2	60.0	958	6	CA487415	AGENCOURT
33	747.6	59.6	849	6	CA487531	AGENCOURT
34	747	59.6	796	8	DR156332	AGENCOURT
35	747	59.6	845	7	CR983967	AGENCOURT
36	746.8	59.6	908	2	BE379425	AGENCOURT
37	742.6	59.2	782	2	BI224248	AGENCOURT
38	742	59.2	871	3	BQ225444	AGENCOURT
39	741.4	59.1	1066	2	BG824323	AGENCOURT
40	741.2	59.1	921	6	CA488935	AGENCOURT
41	737.2	58.8	879	2	BG824387	AGENCOURT
42	729	58.1	749	5	BX344842	AGENCOURT
43	727.6	58.0	857	5	BX374627	AGENCOURT
44	704.4	56.2	800	6	CB529027	AGENCOURT
45	698.8	55.7	767	3	BI753664	AGENCOURT

ALIGNMENTS

RESULT 1
CR607242
LOCUS
DEFINITION
full-length cDNA clone CS0DI067YJ01 of Placenta Cot 25-normalized of Homo sapiens (human).
ACCESSION
CR607242
VERSION
CR607242.1 GI:50488049
KEYWORDS
HTC; CNSLT.cDNA.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1266)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
JOURNAL
Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue
REFERENCE
2 (bases 1 to 1266)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.
FEATURES
Location/Qualifiers
source
1..1266
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI067YJ01"
/issue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN

Query Match 99.9%; Score 1252.4; DB 4; Length 1266;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 481 GACTGGGGAGTACATCTGCTAAATTTGATGGTTGTTACTGTGCACAGTTGGGAAAATTTG 540
Db 486 GACTGGGGAGTACATCTGCTAAATTTGATGGTTGTTACTGTGCACAGTTGGGAAAATTTG 545
QY 541 GCAGATGGTTATAGACACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
Db 546 GCAGATGGTTATAGACACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 605
QY 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGGCCAATTTATACAGAAATCCGA 660
Db 606 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGGCCAATTTATACAGAAATCCGA 665
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCACATTTGATGATTCCTGGAAAGTATAAAG 720
Db 666 CAGTACTGCAATCACTGGCGAAATTTTGTGCACATTTGATGATTCCTGGAAAGTATAAAG 725
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTGCTGGACAGGG 780
Db 726 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTGCTGGACAGGG 785
QY 781 GGTGGAAAGACCCAGATATGTTAGTATTTGGCAACTTTGGCCCTCAGCTGGAATCAGCA 840
Db 786 GGTGGAAAGACCCAGATATGTTAGTATTTGGCAACTTTGGCCCTCAGCTGGAATCAGCA 845
QY 841 GTAACCTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTATTTATGTCCTAATGACCTC 900
Db 846 GTAACCTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTATTTATGTCCTAATGACCTC 905
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Db 906 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT 965
QY 961 CAGAACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db 966 CAGAACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1025
QY 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGATGATTAACCGGAGGAGATTGGT 1080
Db 1026 GAACGACCTCTCTCAGGCTTAGCTGGGCTGATGATTAACCGGAGGAGATTGGT 1085
QY 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTCTTAATCCT 1140
Db 1086 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTCTTAATCCT 1145
QY 1141 GCCTGCTTATACACAGCTCTCCCTGTGAAAGGAGTGGTCTTATGATGAGT 1200
Db 1146 GCCTGCTTATACACAGCTCTCCCTGTGAAAGGAGTGGTCTTATGATGAGT 1205
QY 1201 TCAAGGTTAAGAGTACATATAATCCACAGGCACTGTTTGTCTTCAAG 1248
Db 1206 TCAAGGTTAAGAGTACATATAATCCACAGGCACTGTTTGTCTTCAAG 1253

RESULT 4
LOCUS AY408540 1290 bp DNA linear GSS 15-DEC-2003
DEFINITION Homo sapiens GLA gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY408540
VERSION AY408540.1 GI:39764511
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 1290)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios

JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1290)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1..1290
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>1290
/gene="GLA"
/locus_tag="HCM3258"
ORIGIN
Query Match 92.5%; Score 1160.4; DB 10; Length 1290;
Best Local Similarity 92.6%; Pred. No. 0;
Matches 1161; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAACCCAGNACTACATCTGGGCTGCGCGCTTGGCTTCGCTTCCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGNACTACATCTGGGCTGCGCGCTTGGCTTCGCTTCCTGGCC 60
QY 61 CTGCTTTCTTGGACATCTCTGGGCTTAGAGCACTGGACAAATGGATGGCAAGAGCGCT 120
Db 61 CTGCTTTCTTGGACATCTCTGGGCTTAGAGCACTGGACAAATGGATGGCAAGAGCGCT 120
QY 121 ACCATCGGCTGGCTGCATCTGGAGCGCTTCATGTGCAACCTTGAAGTGGCAAGAGCGCA 180
Db 121 ACCATCGGCTGGCTGCATCTGGAGCGCTTCATGTGCAACCTTGAAGTGGCAAGAGCGCA 180
QY 181 GATTCCTGCATCAGTCAGAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
Db 181 GATTCCTGCATCAGTCAGAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
QY 241 TGGAGGATGTCAGGTTATGATGATACCTCTGCATTGATGACTGTGGTGGCTCCCCAAGA 300
Db 241 TGGAGGATGTCAGGTTATGATGATACCTCTGCATTGATGACTGTGGTGGCTCCCCAAGA 300
QY 301 GATTTCAGAGGAGAGACTTCAGGAGAGCCCTCAGGCGCTTTCCTCATGGGATTCGCGAGCTA 360
Db 301 GATTTCAGAGGAGAGACTTCAGGAGAGCCCTCAGGCGCTTTCCTCATGGGATTCGCGAGCTA 360
QY 361 GCTAATTTATGTTTCCACAGCAAGGAGCTGAGCTAGGAGATTTATGCAGATGTTGAAATATA 420
Db 361 GCTAATTTATGTTTCCACAGCAAGGAGCTGAGCTAGGAGATTTATGCAGATGTTGAAATATA 420
QY 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTTGATGCCAGAGCTTTGCT 480
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTTGATGCCAGAGCTTTGCT 480
QY 481 GACTGGGAGTACATCTGCTAAATTTGATGGTGTTCCTCATGGGATTCGCAAGATTTG 540
Db 481 GACTGGGAGTACATCTGCTAAATTTGATGGTGTTCCTCATGGGATTCGCAAGATTTG 540
QY 541 GCAGATGTTTATAAGACACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
Db 541 GCAGATGTTTATAAGACACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTTATATATGTGGCCCTTTCAAAGGCCAATTTATACAGAAATCCGA 660
Db 601 NNN 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCACATTTGATGATTCCTGGAAAGTATAAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCACATTTGATGATTCCTGGAAAGTATAAAG 720

```
Qy 721 AGTATCTTGAGCTGGAATCTTTTAAACAGAGAGAAATGTTGATGTTGCTGACCAAGG 780
Db 721 AGTATCTTGAGCTGGAATCTTTTAAACAGAGAGAAATGTTGATGTTGCTGACCAAGG 780
Qy 781 GGTGGGAATGACCCAGATATGTTAGTGCATGTTGGCAACTTTGGCCCTCAGCTGGAATCAGCA 840
Db 781 GGTGGGAATGACCCAGATATGTTAGTGCATGTTGGCAACTTTGGCCCTCAGCTGGAATCAGCA 840
Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGTTCTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGTTCTAATGACCTC 900
Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATGAAGACGTAATGCGCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATGAAGACGTAATGCGCATCAAT 960
Qy 961 CAGGACCCCTTGGCAGCAGGATGACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGCAGCAGGATGACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTTAGCTGGGCTGTAGCTATGATAAACCGGACAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTTAGCTGGGCTGTAGCTATGATAAACCGGACAGGAGATTGGT 1080
Qy 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGAGAGTGGCTGTAAATCCCT 1140
Db 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGAGAGTGGCTGTAAATCCCT 1140
Qy 1141 GCCTGCTTCATCAGACGCTCTCCCTGTGAAAGGAGTGGCTTCTATGATGAGT 1200
Db 1141 GCCTGCTTCATCAGACGCTCTCCCTGTGAAAGGAGTGGCTTCTATGATGAGT 1200
Qy 1201 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTGTCTCAGCTATATA 1254
Db 1201 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTGTCTCAGCTATATA 1254

RESULT 5
LOCUS AY408541 1290 bp DNA linear GSS 15-DEC-2003
DEFINITION Pan troglodytes GLA gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY408541
VERSION AY408541.1 GI:39764512
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
REFERENCE 1 (bases 1 to 1290)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Fierriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene tricos
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1290)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Fierriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source Location/Qualifiers
1..1290
/organism="Pan troglodytes"
```

```

/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1290
/genes="GLA"
/locus_tag="HCM3258"

ORIGIN

Query Match 90.9%; Score 1140; DB 10; Length 1290;
Best Local Similarity 91.1%; Pred. No. 0;
Matches 1143; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTGGCGCTTCGCTTCCTCGGCC 60
Db 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTGGCGCTTCGCTTCCTCGGCC 60
Qy 61 CTGCTTTCTTGGGACATCCCTGGGCTTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Db 61 CTCTTTTCTTGGGACATCCCTTGGGCTTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Qy 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
Qy 181 GATTCTTCATCAGTCAGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAGGC 240
Db 181 GATTCTTCATCAGTCAGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAGGC 240
Qy 241 TGAAGGATGCAAGTTATGAGTACCTCTGATTTGATGACTGTTGGAGTGGCTCCCAAGA 300
Db 241 TGGAGGATGCAAGTTATGAGTACCTCTGATTTGATGACTGTTGGAGTGGCTCCCAAGA 300
Qy 301 GATTCAAGAGGAGAGCTTCAGGAGACCTTCAGGCTTCTCATGCGGATTCGCGAGCTA 360
Db 301 GATTCAAGAGGAGAGCTTCAGGAGACCTTCAGGAGACCTTCAGGCTTCTCATGCGGATTCGCGAGCTA 360
Qy 361 GCTAATTATGTTTCCACAGCAAGAGCTGAAGCTAGGAGATTTATGAGATGTTGGAATATA 420
Db 361 NNNNNNNNGTTTCCACAGCAAGAGCTGAAGCTAGGAGATTTATGAGATGTTGGAATATA 420
Qy 421 ACCTGCGCAGGCTTCCTCTGGGAGTTTGGATATCTACGACATGATGCCAGACCTTTGCT 480
Db 421 ACCTGCGCAGGCTTCCTCTGGGAGTTTGGATATCTACGACATGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGAGTAGATCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 481 GACTGGGAGTAGATCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Qy 541 GCAGATGTTTATGAGCAGATGCTCTTGGCCCTGAAATAGGACTGGCAGAGCATTTGTGTAC 600
Db 541 GCAGATGTTTATGAGCAGATGCTCTTGGCCCTGAAATAGGACTGGCAGAGCATTTGTGTAC 600
Qy 601 TCCTGTGAGTGGGCTCTTTATATATGTCGCCCTTTTCAAAAGGCCCAATATACAGAAATCCGA 660
Db 601 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 660
Qy 661 CAGTACTGCAATCAGCTGGGAAATTTTGTCTGCAATTTGATGATTTCTGCGGAAAGTATAAG 720
Db 661 CAGTACTGCAATCAGCTGGGAAATTTTGTCTGCAATTTGATGATTTCTGCGGAAAGTATAAG 720
Qy 721 AGTATCTTGGAGCTGAGCATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTGGACCAAGG 780
Db 721 AGTATCTTGGAGCTGAGCATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTGGACCAAGG 780
Qy 781 GGTGGGAATGACCCAGATATGTTAGTGCATGTTGGCAACTTTGGCCCTCAGCTGGAATCAGCA 840
Db 781 GGTGGGAATGACCCAGATATGTTAGTGCATGTTGGCAACTTTGGCCCTCAGCTGGAATCAGCA 840
Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGTTCTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGTTCTAATGACCTC 900
Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATGAAGACGTAATGCGCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATGAAGACGTAATGCGCATCAAT 960
```

Db	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGGCCATCAAT	960	Db	6	ATGCGAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGC-CTTCGCTTCTCGGCC	64
Qy	961	CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGGAGACAACTTTGAAGTGTGG	1020	Qy	61	CTCGTTTCTCGGACATCCCTGGGGCTAGAGCACTGGACAAATGATTTGGCAAGGACGCT	120
Db	961	CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGNNAGACAACTTTGAAGTGTGG	1020	Db	65	CTCGTTTCTCGGACATCCCTGGGGCTAGAGCACTGGACAAATGATTTGGCAAGGACGCT	124
Qy	1021	GAAAGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAAACCGGACGAGATTGGT	1080	Qy	121	ACCATGGGCTGGCTGCTACTGGGAGCGCTTATGTGCAACCTTGAATGCTCCAGGAAGAGCCA	180
Db	1021	GAAAGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAAACCGGACGAGATTGGT	1080	Db	125	ACCATGGGCTGGCTGCTACTGGGAGCGCTTATGTGCAACCTTGAATGCTCCAGGAAGAGCCA	184
Qy	1081	GGACCTCGCTTTATACCATCGAGTGTGCTTCCCTGGGTAAGAGGAGTGGCTGTAAATCCT	1140	Qy	181	GAATTCGTGATCAGTGAGAACTCTTATGAGAGATGGCAGAGCTCATGTGCTCAGAAGGC	240
Db	1081	GGACCTCGCTTTATACCATCGACTGTGCTTCCCTGGGTAAGAGGAGTGGCTGTAAATCCT	1140	Db	185	GAATTCGTGATCAGTGAGAACTCTTATGAGAGATGGCAGAGCTCATGTGCTCAGAAGGC	244
Qy	1141	GCCTGCTTCATCACAGCTCTCCCTGTGAAAGGAAAGCTAGGTTCTATGAATGGACT	1200	Qy	241	TGGAAGGATGCAAGGTTATGAGTACCTCTGTCATTCATGACTTGTGATGGCTCCCAAGA	300
Db	1141	GCCTGCTTCATCACAGCTCTCCCTGTGAAAGGAAAGCTAGGTTCTATGAATGGACT	1200	Db	245	TGGAAGGATGCAAGGTTATGAGTACCTCTGTCATTCATGACTTGTGATGGCTCCCAAGA	304
Qy	1201	TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTTCCTTCAGCTATAA	1254	Qy	301	GAATTCAGAGGAGAGCTTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCGAGCTA	360
Db	1201	TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTTCCTTCAGCTATAA	1254	Db	305	GAATTCAGAGGAGAGCTTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCGAGCTA	364
RESULT 6				Qy	361	GCTAATTAATGTTACAGCAAAAGGACTGAAGCTAGGGATTTATGCGAGATTTGGAAATAAA	420
AL554978				Db	365	GCTAATTAATGTTACAGCAAAAGGACTGAAGCTAGGGATTTATGCGAGATTTGGAAATAAA	424
LOCUS				Qy	421	ACCTCGGAGGCTTCCCTGGGAGTTTTCGATACACGACATTCATGATGCGCAGACCTTTGCT	480
DEFINITION				Db	425	ACCTCGGAGGCTTCCCTGGGAGTTTTCGATACACGACATTCATGATGCGCAGACCTTTGCT	484
ACCESSION				Qy	481	GACTGGGAGTAGATCTCTCAAAAATTTGATGGTTTGTATCTGTGACAGTTGGGAAAAATTG	540
VERSION				Db	485	GACTGGGAGTAGATCTCTCAAAAATTTGATGGTTTGTATCTGTGACAGTTGGGAAAAATTG	544
KEYWORDS				Qy	541	GCAGATGGTTATAAGCAGATGTCCTTGGCCCTGATAGGACTGCGCAGAGCAATTTGTGTAC	600
SOURCE				Db	545	GCAGATGGTTATAAGCAGATGTCCTTGGCCCTGATAGGACTGCGCAGAGCAATTTGTGTAC	604
ORGANISM				Qy	601	TCCTGTGAGTGCCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA	660
REFERENCE				Db	605	TCCTGTGAGTGCCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA	664
AUTHORS				Qy	661	CAGTACTGCAATCACTGGCGAAAATTTTGTGACATTTGATGATTTCTTGAAAAGTATAAAG	720
TITLE				Db	665	CAGTACTGCAATCACTGGCGAAAATTTTGTGACATTTGATGATTTCTTGAAAAGTATAAAG	724
JOURNAL				Qy	721	AGTATCTTGGACTGAGACATCTTTTAAACAGAGAGAAATTTGTGATTTGCTGGACAGGG	780
COMMENT				Db	725	AGTATCTTGGACTGAGACATCTTTTAAACAGAGAGAAATTTGTGATTTGCTGGACAGGG	784
				Qy	781	GGTTGGAATGACCCAGATATGTTAGTGATTTGCGCAACTTTTGGCCCTCAGCTGGAATCAGCAA	840
				Db	785	GGTTGGAATGACCCAGATATGTTAGTGATTTGCGCAACTTTTGGCCCTCAGCTGGAATCAGCAA	844
FEATURES				Qy	841	GTAACTCAGATGGCCCTCTGGGCTA-TCATGGCTGCTCCTTTATTCATGCTTAATGACCT	899
source				Db	845	GTAACTCAGATGGCCCTCTGGGCTA-TCATGGCTGCTCCTTTATTCATGCTTAATGACCT	904
				Qy	900	CGGACATCAGCCCTCAAGCCAAAGGCTCTCCTTCAGATTAAGGACCTTAATTCGCATCAA	959
				Db	905	CGGACATCAGCCCTCAAGCCAAAGGCTCTCCTTCAGATTAAGGACCTTAATTCGCATCAA	964
				Qy	960	TCAGGACCCCTTGGGCAAGGAGGCTTACAGCTTAGACAGGAGAGCAACTTTTGAAGTGTG	1019
				Db	965	TCAGGACCCCTTGGGCAAGGAGGCTTACAGCTTAGACAGGAGAGCAACTTTTGAAGTGTG	1022
ORIGIN				Qy	1020	GGAAACGACCTCTC 1032	
				Db	1023	GGNMGACTCTCTC 1035	
Query Match							
Best Local Similarity							
Matches 1012; Conservative							
1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGCCTTCCTCGGCC	60					
RESULT 7							

QY	841	GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTATTTCATGTCTAATGACCTC	900
DB	846	GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCT-CYTTATTTCATGTCTAATGACCTC	904
QY	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCGCATCAAT	960
DB	905	CGACACATCAGCCCTCAAGCC-AAAGCTCTCTTCAGGATAAGGACGTAATTCG-CATCAAT	962
QY	961	CAGGACCCCTTGGGCAAGCAAGGG	984
DB	963	CAGGA-CCCTTGGGCAAGCAAGGK	985
RESULT 9			
BM564282			
LOCUS			
DEFINITION			
5', mRNA Sequence.			
BM564282			
BM564282.1 GI:18811955			
EST.			
Homo sapiens (human)			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;			
Hominidae; Homo.			
1 (bases 1 to 1067)			
NIH-MGC http://mgc.nci.nih.gov/.			
National Institutes of Health, Mammalian Gene Collection (MGC)			
Unpublished (1999)			
Contact: Robert Strausberg, Ph.D.			
Email: cgaabs-r@mail.nih.gov			
Tissue Procurement: Life Technologies, Inc.			
cDNA Library Preparation: Life Technologies, Inc.			
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
DNA Sequencing by: Agencourt Bioscience Corporation			
Clone distribution: MGC clone distribution information can be			
found through the I.M.A.G.E. Consortium/LLNL at:			
http://image.llnl.gov			
Plate: LLMI2759 row: p column: 07			
High quality sequence stop: 761.			
FEATURES			
source			
1..1067			
Location/Qualifiers			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/clone="IMAGE:5742198"			
/tissue_type="medulla"			
/lab_host="DH10B"			
/clone_lib="NIH_MGC_119"			
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;			
Site 2: EcoRV (destroyed); RNA source normal medulla from			
anonymous male age 27. Library is oligo-dT primed and			
directionally cloned (EcoRV site is destroyed upon			
cloning). Average insert size 1.3 kb, insert size range			
0.9-3 Kb. Library is normalized and enriched for			
full-length clones and was constructed by C. Gruber			
(Invitrogen). Research Genetics tracking code 013. Note:			
this is a NIH_MGC Library."			
ORIGIN			
Query Match 73.0%; Score 915.8; DB 3; Length 1067;			
Best Local Similarity 96.4%; Pred. No. 5.6e-263;			
Matches 1009; Conservative 0; Mismatches 31; Indels 7; Gaps 7;			
QY	6	GCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCGCTTCGCTTCCTGGCCCTCGT	65
DB	15	GCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCGCTTCGCTTCCTGGCCCTCGT	74
QY	66	TTCTCTGGGACATCCCTTGGGGCTAGAGCACTGGCAATGGATTGGCAGGAGCGCTACCAT	125
DB	75	TTCTCTGGGACATCCCTTGGGGCTAGAGCACTGGCAATGGATTGGCAGGAGCGCTACCAT	134

```
Qy 126 GGGCTGGCTGACCTGGAGGCGCTTCATGTGCAACCTTGACCTGCCAGGAAGCCAGATTC 185
Db 135 GGGCTGGCTGACCTGGAGGCGCTTCATGTGCAACCTTGACCTGCCAGGAAGCCAGATTC 194
Qy 186 CTGCATCAGTGAAGACTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAAGCTGGAA 245
Db 195 CTGCATCAGTGAAGACTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAAGCTGGAA 254
Qy 246 GGATGCAAGGTTATCAGTACCTCTGCATTTGATGACTGTTGGATGGCTCCCAAGAGATTC 305
Db 255 GGATGCAAGGTTATCAGTACCTCTGCATTTGATGACTGTTGGATGGCTCCCAAGAGATTC 314
Qy 306 AGAAGGAGACTTCAGGAGACCTCTAGCGCTTTCTCATGGGANTTCGCGAGCTAGCTAA 365
Db 315 AGAAGGAGACTTCAGGAGACCTCTAGCGCTTTCTCATGGGANTTCGCGAGCTAGCTAA 374
Qy 366 TTATGTTACAGCAAGGACTGAGCTAGGATTTATGCGAGTGTGGAAATTAAGACTG 425
Db 375 TTATGTTACAGCAAGGACTGAGCTAGGATTTATGCGAGTGTGGAAATTAAGACTG 434
Qy 426 CGCAGGCTCTCCCTGGAGTTTGGATACTACGACATTTGATGCCCGAGACCTTTGCTGACTG 485
Db 435 CGCAGGCTCTCCCTGGAGTTTGGATACTACGACATTTGATGCCCGAGACCTTTGCTGACTG 494
Qy 486 GGGAGTAGATCTGCTAAAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTGGCAGA 545
Db 495 GGGAGTAGATCTGCTAAAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTGGCAGA 554
Qy 546 TGGTTATAGCAGATGCTCTGGGCTTGAATAGAGCTGGCAGAGCAATTTGCTACTCTG 605
Db 555 TGGTTATAGCAGATGCTCTGGGCTTGAATAGAGCTGGCAGAGCAATTTGCTACTCTG 614
Qy 606 TGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATATACAGAAATCCGACAGTA 665
Db 615 TGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATATACAGAAATCCGACAGTA 674
Qy 666 CTGCAATCACTGGCGAAATTTGCTGATGATGATTCCTGGAAAAAGTATAAGAGTAT 725
Db 675 CTGCAATCACTGGCGAAATTTGCTGATGATGATTCCTGGAAAAAGTATAAGAGTAT 734
Qy 726 CTTGGAGTGCATCTTTTAAACGAGAGAGATTTGTTGATGTTGCTGGACCA -GGGGGTT 784
Db 735 CTTGGAGTGCATCTTTTAA -CAGGAGAGATTTGTTGATGTTGCTGGACGAGGGGGTT 793
Qy 785 GGAATGACCCAGATATGTTAGTGTGATGCGCAACTTTGGCCCTCAGCTGGAATCAGCAAGTAA 844
Db 794 GGAATGACCCAGATATGTTAGTGTGATGCGCAACTTTGGCCCTCAGCTGGAATCANGCAGTAA 853
Qy 845 CTCAGATGGCCCTCTGGGCTATCATGGCTGCT -CCTTTATTTCATGCTTAATGACCTCCGA 903
Db 854 CTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCCTTTATTTCATGCTTAATGACCTCCG 913
Qy 904 CACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAGGACGTAATGCGATCAATCAG 963
Db 914 AACATCAACCTC -AGCCAAAGCTCTCTTTCAAGATAGGACGT -ATTGCTCATCCATCAG 971
Qy 964 GACCCCTTGGGCAAGCAAGGTA -CCAGCTTAGACAGGAGACAACCTTTCAAGTGTGGGA 1022
Db 972 GAACCTTGGNGCCAGCAGGATACCCNNCTTAGACCGGAAAAACAACCTTTAAATGTGGGA 1031
Qy 1023 ACG -ACCTCTCTCAGGCTTAGCCCTGGG 1048
Db 1032 ACGNACTCTCTCAGGCTTTAAACCGGG 1058
```

```
RESULT 10
AL577581/c
LOCUS
DEFINITION
AL577581 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
cDNA clone CS0DK009YE12 3-PRIME, mRNA sequence.
ACCESSION
AL577581
VERSION
AL577581.3 GI:46256603
```

KEYWORDS

SOURCE
ORGANISM

EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.

REFERENCE
1 (bases 1 to 1026)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 16, 2001 this sequence version replaced gi:31315840.

CONTACT: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
10506.r

For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0DK009BC06NPl&c=10506.r.

Location/Qualifiers
1..1026
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK009YE12"
/cell_type="HELA"
/clone_lib="HELA"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES
source

Query Match 73.0%; Score 915; DB 1; Length 1026;
Best Local Similarity 98.3%; Pred. No. 9.5e-263;
Matches 939; Conservative 4; Mismatches 10; Indels 2; Gaps 2;

289 GCTCCCCAAGAGATTCAGAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTTCTCATGGG 348
Db 955 GCTCCCCAAGAGATTCAGAA -GCAGACTTCA -GCAGAGCCCTCAGCGCTTTCTCATGGG 898
Qy 349 ATTGCGCAGCTAGCTAATATGTTTACAGCAAGGAGCTGAAGCTAGGATTTATGACAGAT 408
Db 897 ATTGCGCAGCTAGCTAATATGTTTACAGCAAGGAGCTGAAGCTAGGATTTATGACAGAT 838
Qy 409 GTTGGAAATAAAACCTGCGAGGCTTCCCTGGAGATTTGGATCTACGACATTTGATGCC 468
Db 837 GTTGGAAATAAAACCTGCGAGGCTTCCCTGGAGATTTGGATCTACGACATTTGATGCC 778
Qy 469 CAGACCTTTGCTGACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTACTGTGACAGT 528
Db 777 CAGACCTTTGCTGACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTACTGTGACAGT 718
Qy 529 TTGGAATAATTTGCGAGATGTTATAGCAGATCTCTTGGCCCTGGAATAGGAGTGGCAGA 588
Db 717 TTGGAATAATTTGCGAGATGTTATAGCAGATCTCTTGGCCCTGGAATAGGAGTGGCAGA 658
Qy 589 AGCATTTGTGACTCTCTGAGTGGCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATTTAT 648
Db 657 AGCATTTGTGACTCTCTGAGTGGCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATTTAT 598
Qy 649 ACAGAAATCCGACAGTACTGCAATCACTGCGGAAATTTTCTGCTGACATTTGATGTTCTGG 708
Db 597 ACAGAAATCCGACAGTACTGCAATCACTGCGGAAATTTTCTGCTGACATTTGATGTTCTGG 538
Qy 709 AAAAGTATTAAGAGTATCTTGGACTGACATCTTTTAAACCAGGAGAGAAATTTGTTGATGTT 768
Db 537 AAAAGTATTAAGAGTATCTTGGACTGACATCTTTTAAACCAGGAGAGAAATTTGTTGATGTT 478

				AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	JOURNAL	Nature	409, 685-690 (2001)	REFERENCE	
AUTHORS	AUTHORS				5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs			
JOURNAL	JOURNAL	Nature	420, 563-573 (2002)	REFERENCE	
AUTHORS	AUTHORS				6 (bases 1 to 1376) Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kohji,M., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Ohaato,N., Okazaki,Y., Saito,R., Saichou,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,P., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216]
				COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.jp/ URL:http://fantom.gsc.riken.jp/ Location/Qualifiers 1..1376 /organism="Mus musculus" /mol_type="mRNA" /strain="CS7BL/6J" /db_xref="FANTOM DB:E330039P08" /clone="E330039P08" /sex="female" /tissue_type="ovary" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="2 days pregnant adult"
	CDS				21..1286 /note="unnamed protein product; galactosidase, alpha (MGDI[MGI:1347344, GB NM_013463, evidence: BLASTN, 99%, match=1373) putative" /codon_start=1 /protein_id="BAC35819.1" /db_xref="GI:26344325"
					/translation="MAWKLRDRTLVCELCALCPALVFWSILGVRLDNLCAFTPTM GWLWHERFMCLNDQOEPCDACISQLFMQAEMLMVSDGRADGYDICIDCWNPAPER DSKGLRDLQPFPSPGISKHLANYHSGKLGIYADVGNKTCCAGPFSGSYDIDAQT PADWGVLKKDPGCDCSDSWLENGKYMALALTGRSIVYSWEFLYLRFPHKNY TDIOYYCHNRNFDDVDYSWESIKNLSWTVOKEIVEVAGPSWDPLMGVCFR KENHQVVCTQMALWAAPALMSLDLROISSOAKLLQNKDVIANDPLGKGCVCFR KENHWERWERPLUSLNAAVAVRNQIEIGGPCPYTIQISSLOGLACNPGCILITOLLEPR KVHGLFWETLTLTRVPNSGTGTVLFRLER" 1358..1363 /note="putative" 1376 /note="putative"
					polya_signal polya_site ORIGIN

Query Match	71.5%;	Score	897.2;	DB	4;	Length	1376;	
Best Local Similarity	82.2%;	Pred. No.	2.4e-257;					
Matches 1031;	Conservative	0;	Mismatches	223;	Indels	0;	Gaps	0;
QY	1	ATGCAGCTGAGGAAACCCAGAACTACATCTGGCGCTGCGCGCTTCGGCTTCCTCGTCCG	60					
Db	27	ATGAGCTTTTGACGAGAGATACGGCGCTGGTCTGTGAGCTTGGCGCTTTGTCCCTGGCT	86					
QY	61	CTCGTTTCCTGGGACATCCCTGGGGCTPAGAGCACTGGCAATGATGGCAAGGACGCTT	120					
Db	87	TTAGTTTCTTGGAGCACTTCTGGGGCTCAGAGCATTTGGCAATGGCTTTGGCGCGGACTCCT	146					
QY	121	ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTCCGAGGAGAGCCA	180					
Db	147	ACTATGGGCTGGCTGGCAATGGGAAACGTTTCATGTGCAACCTTGACTGCCAAGAGAGCCT	206					
QY	181	GATTTCCTGCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC	240					
Db	207	GATGCTGCTGATAGTGAGCACTGTTTCATGCGAGATGGCAGAGCTCATGGTCTCTGATGC	266					
QY	241	TGGAAGGATCGAGGTATAGTACCTCTGTCATTGATGACTGTTGGATGGCTCCCAAAGA	300					
Db	267	TGGCGGATCGAGGTATAGTACCTCTGTCATTGATGACTGTTGGATGGCTCCCGAGAGG	326					
QY	301	GATTTCAGAGCGCAGACTTCAGGCGAGACCCCTCAGCGCTTTCCTCATGGATTCGCCAGCTA	360					
Db	327	GATTTCAGAGCGCAGCTTCAGGCGAGATCCCAACGCTTTCCTCATGGATTCACCAACCTC	386					
QY	361	GCTAAATATCTTCACAGCAAGAGCTGAAGCTAGGAGTTTATGCGAGATGTTGGAATAAA	420					
Db	387	GCAATTTACGTCACAGCAAGAGATGGAAGCTAGGAGTTTATGCGAGATGTTGGAATAAA	446					
QY	421	ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATGATGCCAGACCTTTGCT	480					
Db	447	ACCTGTGCAGTTTCCCGGAGTTTGGATCCTATGACATGATGCGCAGACATTTGCT	506					
QY	481	GACTGGGAGTAGATCTGCTAAATTTGATGGTTTATGTCGACAGTGTGGAAATTTG	540					
Db	507	GACTGGGCGCTAGATCTGCTAAATTTGATGGTTTGTCACTGACAGTGTAGTATCCTTG	566					
QY	541	GCAGATGTTTATAGCACATGCTCTGGCCCTCAATAGGAGCTGCGAGAGCATTTGTATC	600					
Db	567	GAGATGTTTATAGTACATGCGCTTGGCCCTTGAACAGGACAGCGCGAAGCATTTGATAC	626					
QY	601	TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGCCCAATTTATACAGAAATCCGA	660					
Db	627	TCCTGTGAGTGGCCACTTTATTTGAGACCCCTTTCAATAAGCCCAATTTATACAGATATCAA	686					
QY	661	CAGTACTGCATCACTGGCGAAATTTTGTGACATGATGATTCCTGGGAAAGATATAAG	720					
Db	687	TATTACTGCAATCAITGGAGAAATTTTGTATGATGTTTATGATTCCTGGGAAAGCATAAAG	746					
QY	721	AGTATCTTGACCTGGACATCTTTTACGAGGAGAGATTTGTCATGTTGCTGACACAGG	780					
Db	747	AATATCTTGTCTTGGACAGTGGTTTACGAGAGAGATTTGTAAGTCTGCTGGACAGGC	806					
QY	781	GGTTGGAATGACCCAGATATGTTAGTATGGCAACTTTTGGCCCTCAGCTGGATCAGCAA	840					
Db	807	AGCTGGATGACCCAGACATGTTAGTATGCGCAACTTTGGCCCTCAGTTGGGACAGCAG	866					
QY	841	GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTAATGACCTC	900					
Db	867	GTGACACAGATGGCCCTCTGGGCTATCATGGCGCTCCCTACTCATGTCTCAACGATCTG	926					
QY	901	CGACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGTAAGGACCTTAATGCGATCAAT	960					
Db	927	CGCAAAATCAGCTCTCAAGCCAAAGCTCTGCTTTCAGATAAGGATGTAATTTGCCATCAAC	986					
QY	961	CAGGACCCCTTGGGCAAGCAAGGATACAGCTTTAGACAGGAGCAACAATTTGAAGTGTGG	1020					
Db	987	CAAGACCCCTTGGGCAAGCAAGGCTACTGTTTCAGAAAGGAACCAACATTTGAGTTTGG	1046					
QY	1021	GAACGACCTCTCTCAGGCTTTAGCTGCTGGCTGTAGCTATGATATAAACCGGAGGAGTTGCT	1080					

Db	1047	GAACGGCCACTCTCCAAACCTAGCCTGGCTGTGGCTGTGAGAAACCTTCGAGGAGATTGGT	1106					
QY	1081	GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGTAAAGAGAGTGGCTGTAAATCCT	1140					
Db	1107	GGACCTTGCTCTTATACCATCCAGATTTCTTCCCTGGTAGAGGACTAGCCTGCAATCCT	1166					
QY	1141	GCTGCTTCATCAACAGCTCTCCCTCTGTGAAAGGAAGCTAGGTTCTATGAATGACT	1200					
Db	1167	GGCTGCATCACTACTCAGCTTCTCCCGAGAAAGTACACCTAGGCTTCTATGAATGACT	1226					
QY	1201	TCAGGTTAAGAGTACACATAAATCCACAGGCACTGTTTTCCTCAGCTATAA	1254					
Db	1227	TTGACCTTAAAACTCGAGTAAACCCCTCAGGCACTGTTTGTTCGGTTAGAA	1280					

RESULT 12	
AK040081	
LOCUS	2962 bp mRNA linear HTC 03-APR-2004
DEFINITION	Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched library, clone:A430057F16 product:galactosidase, alpha, full insert sequence.
ACCESSION	AK040081
VERSION	1 GI:263333580
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	1 Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999) 10349636
REFERENCE	2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) 11042159
REFERENCE	3 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsumura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000) 11076861
REFERENCE	4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)
REFERENCE	5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 6 (bases 1 to 2962)
REFERENCE	Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,


```
C0645623.1 GI:50567117  
EST.  
KEYWORDS Macaca mulatta (rhesus monkey)  
SOURCE Macaca mulatta  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Cercopithecoidea; Cercopithecidae; Macaca.  
  
REFERENCE 1 (bases 1 to 922)  
AUTHORS Magness,C.L., Fellin,P.C., Thomas,M.J., Korth,M.J., Agy,M.B.,  
Prohl,S.G., Fitzgibbon,M., Scherer,C.A., Miner,D.G., Katze,M.G. and  
Iadonato,S.P.  
TITLE Analysis of the Macaca mulatta transcriptome and the sequence  
divergence between Macaca and human  
JOURNAL Genome Biol. 6 (7), R60 (2005)  
PUBMED 1598449  
COMMENT Contact: C. Magness  
Illuminigen Biosciences Inc.  
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA  
Tel: 2063780400  
Fax: 2063780408  
Email: cmagness@illumigen.com  
Sequenced on 2004.03.02. 796 Q20 bases. Library Preparation: Prof.  
Michael Katze Lab at University of Washington DNA Sequencing:  
Illuminigen Biosciences Inc. For further information, see  
http://www.macaque.org
```

```
PCR Primers  
FORWARD: CCCTCACTAAAGGGACAAA  
BACKWARD: CACTATAGGCGAATTGGTGA  
Insert Length: 922 Std Error: 0.00  
Plate: CJ000111 row: B column: 08  
Seq primer: CCCTCACTAAAGGGACAAAA  
POLYA=No.
```

```
FEATURES             Location/Qualifiers  
     ..+..            1..922  
                     /organism="Macaca mulatta"  
                     /mol_type="mRNA"  
                     /strain="Indian"  
                     /db_xref="taxon:9544"  
                     /clone="IBIW:23923"  
                     /sex="female"  
                     /dev_stage="adult"  
                     /lab_host="Electromax DH10B"  
                     /clone_lib="Katze_WMLV"  
                     /note="Organ: liver; Vector: pDONR 222; Site 1: Berg I;  
Site 2: Berg I; Created from CloneMiner cDNA Library  
Construction kit (catalog #18249-029)"
```

ORIGIN	CONSTRUCTION KIL (Catalog #18249-029)	Query Match	70.4%	Score 883.2	DB 7	Length 922	Indels 0	Gaps 0
		Best Local Similarity	98.0%	Pred. No. 3.2e-253				
		Matches 894	Conservative 0	Mismatches 18				
QY	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGTGGCGCTTGGCTTCGCTTCTCTGGCC	60					
DB	11	ATGCAGCTGAGGAACCCAGAAATGCATCTGGGTGGCGCTTGGCTTCTCTCTCTGGCC	70					
QY	61	CTCGTTTCCTGGGACATCCCTGGGGCTAGACACTGGACAAATGATGGCAAGGAGCGCT	120					
DB	71	CTCGTTTCCTGGGACATCCCTGGGGCCAGAGCACTGGACAATGGATGGCCAGGAGCGCT	130					
QY	121	ACCATGGGCTGGCTGCACCTGGGAGCGCTTATGTGCAACCTTGACTGCCAGGAAGAGCCA	180					
DB	131	ACCATGGGCTGGCTGCACCTGGGAGCGCTTATGTGCAACCTTGACTGCCAGGAAGAGCCA	190					
QY	181	GATTCTTCGATCAGTGAAGAGCTTTCATGTGAGATGGCAGAGCTCATGTGTCAGAAAGGC	240					
DB	191	GATTCTTCGATCAGTGAAGAGCTTTCATGTGAGATGGCAGAGCTCATGTGTCAGATGGC	250					
QY	241	TGGAGGATCGAGGTATAGTACCTCTGCATGTATGACACTGTTGGATGGCTCCCCAAAGA	300					
DB	251	TGGAAGGATCGAGGTATAGTACCTCTGCATGTATGACACTGTTGGATGGCTCCCCAAAGA	310					
QY	301	GATTTCAGGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA	360					

Michael Katze Lab at University of Washington DNA Sequencing:
 Illumigen Biosciences Inc. For further information, see
<http://www.macaque.org>
 PCR Primers

FORWARD: CCTCTACTAAAGGGAACAAA
 BACKWARD: CACTATAGGCGAATGGTA
 Insert length: 1133 Std Error: 0.00
 Plate: CL000111 row: E column: 08
 Seq primer: CCCTCACTAAAGGGAACAAA
 POLYA=No.

FEATURES

Location/Qualifiers

source

1..1133
 /organism="Macaca mulatta"
 /mol_type="mRNA"
 /strain="Indian"
 /db_xref="taxon:9544"
 /clone="IBIUM:23923"
 /sex="female"
 /dev_stage="adult"
 /lab_host="Electromax DH10B"
 /clone_lib="Katze MMLV"
 /note="Organ: liver; Vector: pDONR 222; Site 1: BerG I;
 Site 2: BerG I; Created from CloneMiner cDNA Library
 Construction kit (catalog #18249-029)"

ORIGIN

Query Match 70.3%; Score 881.8; DB 7; Length 1133;
 Best Local Similarity 94.6%; Pred. No. 9.3e-253;
 Matches 924; Conservative 0; Mismatches 52; Indels 1; Gaps 1;

QY 264 CCTCTGATGATGACTGTGGATGCTCCCAAGAGATTACAGAGGAGCTTCAGGC 323
 DB 1041 CCTCTGCCCCGGGACTTTTGGAGATTCCCAAGAGACTCGAAGCGCCCTCGGC 982

QY 324 AGACCTCAGCGCTTCTCATGGATTGCCAGCTAGCTAAATTATGTTACACAGCAAGG 383
 DB 981 GGGCGTTAGTTTCTTCATGGATTGCCAGTTAGCTAATTATGATACAGCAAGG 922

QY 384 ACTGAAGCTAGGATTTATCAGATGTTGAAATAAAACCTGCGCAGGCTTCCTCGGAG 443
 DB 921 AATGAAATAGGATTTATGACAGCTTGGAAATAAAACCTGCGCAGGCTTCCTCGGAG 862

QY 444 TTTTGGATACAGACATTGATGCCAGACTTGTCTGACTGGGGAGTAGACTGCTATAA 503
 DB 861 TTTTGGATACATGACATTGATGCCAGTCTTGTGCTGACTGGGGAGTAGATGTTGAA 802

QY 504 ATTGTAGTTGTTACTGTGACAGTTTGGAAAATTTGGCAGATGTTTAAAGCACATGTC 563
 DB 801 ATTGTAGTTGTTACTGTGACAGTTTGGAAAATTTGGCAGATGTTTAAAGCACATGTC 742

QY 564 CTTGGCCCTGAATAGGACTGGCAGACATGTTGTAATCTCTGTGAGTGGCTCTTTATAT 623
 DB 741 GTTGGCCCTGAATAGGACTGGCAGACATGTTGTAATCTCTGTGAGTGGCTCTTTATAT 682

QY 624 GTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGACAGTACTGCAATCACTGGCCAAA 683
 DB 681 GTGGCCATTTCAAAGCCCAATTATACAGAAATCCGACAGTACTGCAATCACTGGCCAAA 622

QY 684 TTTTGTGACATTCATGATTCCTGAAAAGATATAAGAGATATCTTGGACTGGACATCTTT 743
 DB 621 TTTTGTGACATTCATGATTCCTGAAAAGATATAAGAGATATCTTGGACTGGACATCTTT 562

QY 744 TAACCAAGAGAGAAATTTGATGTTGTGACACAGGGGGTTGGAATGACCCAGATATGTT 803
 DB 561 TAACCAAGAGAGAAATTTGATGTTGTGACACAGGGGGTTGGAATGACCCAGATATGTT 502

QY 804 AGTCAATGGCACTTTGGCCCTCAGCTGGAATCAGCAAGTAACTCAGATGGCCCTCGGC 863
 DB 501 AGTCAATGGCACTTTGGCCCTCAGCTGGAATCAGCAAGTAACTCAGATGGCCCTCGGC 442

QY 864 TATCATGGCTGCTCTTTTATTCATGCTAATGACTCCGACACATCAGCCCTCAAGCAA 923
 DB 441 TATCATGGCTGCTCTTTTATTCATGCTAATGACTCCGACACATCAGCCCTCAAGCAA 382

QY 924 AGCTCTCCTTCAGGATAAGGACGTAATTGCCATCAATCAGGACCCCTTGGGCAAGCAAGG 983
 DB 381 AGCTCTCCTTCAGGATAAGGATGTAATTGCCATCAATCAGGACCCCTTGGGCAAGCAAGG 322

QY 984 GTACCAAGCTTTAGACAGGGAGACAACTTTGAAGTGTGGGAACGACCTCTCTCAGGCTTAGC 1043
 DB 321 GTACCAAGCTTTAGACAGGGAGACAACTTTGAAGTGTGGGAACGACCTCTCTCAGGCTTAGC 262

QY 1044 CTGGGCTGTAGCTATGATATAAACCAGGAGAGATTGGTGGACCTCGCTCTTATACCATCGC 1103
 DB 261 CTGGGCTGTAGCTATGATATAAACCAGGAGAGATCGGTGGACCTCGCTCTTATACCATCGC 202

QY 1104 AGTTGCTTCCCTGGGTAAAGGAGTGCCTGTAATCTCTGCTGCTTATCATCACAGGCTCTCT 1163
 DB 201 GGTGTCTTCCCTGGGTAAAGGAGTGCCTGTAATCTCTGCTGCTTATCATCACAGGCTCTCT 142

QY 1164 CCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACTTCAAGGTTAAGAAAGTCACTATAA 1223
 DB 141 CCCTG-GAAAAGGAAGCTAGGGTTCTATGAATGGACTTCAAGGTTAAGAAAGTCACTATAA 83

QY 1224 TCCACAGGCACTGTTT 1240
 DB 82 TCCACAGGCACTGTTT 66

RESULT 15

AL575861/c

LOCUS

DEFINITION

AL575861 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA

clone CS0D1067YJ01 3-PRIME, mRNA sequence.

ACCESSION

AL575861

VERSION

AL575861.3 GI:46248771

SOURCE

EST.

KEYWORDS

Homo sapiens (human)

ORGANISM

Homo sapiens

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominiidae; Homo.

REFERENCE

1. (bases 1 to 1071)

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

On Feb 16, 2001 this sequence version replaced gi:31314157.

COMMENT

Contact: Genoscope Centre National de Sequencage

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

10506.r

For more information about this cluster, see

<http://www.genoscope.cns.fr/cdna?s=CS0D1067CE01NFI&c=10506.r>.

Location/Qualifiers

1..1071

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0D1067YJ01"

/tissue_type="PLACENTA COT 25-NORMALIZED"

/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 69.8%; Score 874.8; DB 1; Length 1071;
 Best Local Similarity 91.6%; Pred. No. 1.1e-250;
 Matches 979; Conservative 20; Mismatches 61; Indels 9; Gaps 8;

Search completed: December 31, 2005, 21:00:26
Job time : 5621 secs

```
QY 191 TCAGTGAAGACTCTTCATGGAGATGCGAGAGCTCATGGTCTCAGAA-GGCTGGAAGGAT 249
Db |||||
1071 TCAGTGAAGAGCTC-TCATGNANATGNCAGANCTCATGNTCTCAGAACGCTGAAGNAT 1013
QY 250 GCAGGTTATGAGTACCTCTGC--ATTGATGACTGTTGGATGGCT-CCCCAAGAGATTCA 306
Db |||||
1012 GCAGTNTATGAGTACCTCTGCAATTTGATGACTGTTGNATGNTCCCCCAAGAGATNCA 953
QY 307 GAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTAGCTAAT 366
Db |||||
952 GAAGNCACATNCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATNCGCCAGCTAGCTAAT 893
QY 367 TATGTTACAGCAAGCACTGAAGCTAGGAGTTTATGCAATGTTGGAAATAAAACCTGC 426
Db |||||
892 TATGCTCACAGCAAGCACTGAAGCTA-GGATTTATGCAGATGTT-GAATAAAACCTGC 835
QY 427 GCAGGCTTCCTGGGAGTTTGGATCTAGCAATTTGATGCCAGACCTTTGCTGACTGG 486
Db |||||
834 GCAGGCNTCCTGGGAGTTTGGATCTAGCAATTTGATGCCAGACCTTTGCTAACTGG 775
QY 487 GGAGTAGATCTGCTAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAAATTTGGCAGAT 546
Db |||||
774 GGAGTAGATCTGCTAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAAATCTVGCAGAT 715
QY 547 GGTATTAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAAAGCAATTTGTACTCCTGT 606
Db |||||
714 GATWATAAGCACATGTCCTTGGCCCTGAATAGAACTAGCAGAAAGCAATTTGTGTACACCTAT 655
QY 607 GAGTGGCTCTTTATATGTCGCTTTCAAAAGCCCAAT-TATACAGAAATCCGACAGTA 665
Db |||||
654 VAGTRACCTCACTATATCTGCCCTTMYCAAAAGMCCAAATWTATACAAAAATCCGACAGTA 595
QY 666 CTGCAATCACTGGCGAAATTTGCTGACATTTGATGATTTCTGGAAGATATAAGAGTAT 725
Db |||||
594 CTGCAATMACTGGCAAAATTTAGCTGACATTTGATGATTTCTGAAATAAGTATAAGAGTAT 535
QY 726 CTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGATGTTGCTGGACCAGGGGTTG 785
Db |||||
534 CTTGGACTGGACATCTTTTAAACAGAGAAAGAAATTTGATGTTGCTGGACCAGGGGTTG 475
QY 786 GAATGACCCAGATATGTTAGTAGTTGGCACTTTGGCTCAGCTGGNAATCAGCAAGTAAC 845
Db |||||
474 GAATGACCCAGATATGTTAGTAGTTGGCACTTTGGCTCAGCTGGNAATCAGCAASTMAC 415
QY 846 TCAGATGGCCCTCTGGGCTATCATGGCTGCTCCCTTTATTCATGCTCTAATGACCTCCGACA 905
Db |||||
414 TCAGAT-GCCCTCTGGGCTATCATGGCTGCTCCCTTTATTCATGCTCTAATGACCTCCGACA 356
QY 906 CATCAGCCCTCAAGCCAAAGCTCTCCTTTCAGGATAAGGAGCTAATTTGCCATCAATCAGGA 965
Db |||||
355 CATCAGCCCTCAAGCCAAAGCTCTCCTTTCAGGATAAGGAGCTAATTTGCCATCAATCAGGA 296
QY 966 CCCCTTGGGCAAGCAGGGTACAGCTTAGACAGGGAGACAACCTTTGAAGTGGGAAAG 1025
Db |||||
295 CCCCTTGGGCAAGCAGGGTACCRGCTTAGACAGGGAGACAACCTTTGAAGTGGGAAAG 236
QY 1026 ACCTCTCTCAGGCTTAGCCTGGGCTGAGCTATGATAAAACCGGCAGGAGATTGCTGACC 1085
Db |||||
235 ACYCTCTCAGGCTTAGCCTGGGCTGAGCTATGATAAAACCGGCAGGAGATTGCTGACC 176
QY 1086 TCCTCTTTATACCATGCGAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCTGCCTG 1145
Db |||||
175 TCCTCTTTATACCATGCGAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCTGCCTG 116
QY 1146 CTTTCATCACAGCTCCTCCTGTGAAAAGAAAGCTAGGGTTCTATGAATGGAATTTCAAG 1205
Db |||||
115 CTTTCATCACAGCACCCTCCNKGTGAAAAGAAAGCTAGGGTTCTATGAATGGACMTCAAG 56
QY 1206 GTTAAGAGTCAATAAATCCACAGGCACTGTTTTCCTCAGCTATAA 1254
Db |||||
55 GTTAAGAGTCACTWAAATCCCNACAGGCACTGTTTTCCTCAGCTAGAA 7
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 31, 2005, 17:12:51 ; Search time 261 Seconds
(without alignments)
8540.469 Million cell updates/sec

Title: US-10-602-220-15
Perfect score: 1254
Sequence: 1 atgcagctgaggaccagca.....ctgtttgttcagctataa 1254

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues
Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/1_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq.*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq.*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1254	100.0	1254	3	US-09-993-059-15
2	1254	100.0	1254	3	US-10-103-327-15
3	1252.4	99.9	1266	3	US-09-993-059-11
4	1252.4	99.9	1266	3	US-10-103-327-11
5	1252.4	99.9	1278	3	US-09-993-059-7
6	1252.4	99.9	1278	3	US-10-103-327-7
7	1252.4	99.9	1284	3	US-09-993-059-13
8	1252.4	99.9	1284	3	US-10-103-327-13
9	1252.4	99.9	1290	3	US-09-993-059-3
10	1252.4	99.9	1290	3	US-10-103-327-3
11	1252.4	99.9	1296	3	US-09-993-059-9
12	1252.4	99.9	1308	3	US-10-103-327-9
13	1252.4	99.9	1308	3	US-09-993-059-5
14	1252.4	99.9	1308	3	US-10-103-327-5
15	1252.4	99.9	1343	3	US-08-928-881-18
16	1252.4	99.9	1343	3	US-09-543-921-18
17	1252.4	99.9	1343	3	US-09-266-014-3
18	1252.4	99.9	1343	3	US-09-491-759-18
19	1252.4	99.9	1393	2	US-07-602-824A-1
20	1252.4	99.9	1393	2	US-07-983-451-1
21	1252.4	99.9	1393	2	US-08-261-577-6
22	1252	99.8	1272	3	US-09-993-059-17
23	1252	99.8	1272	3	US-10-103-327-17
24	1252	99.8	8234	3	US-09-626-127-14

25	1252	99.8	8234	3	US-09-993-059-34	Sequence 34, Appl
26	1252	99.8	8234	3	US-10-103-327-34	Sequence 34, Appl
27	1217.6	97.1	1304	9	5179023-3	Patent No. 5179023
28	1212.6	96.7	1233	3	US-09-993-059-21	Sequence 21, Appl
29	1212.6	96.7	1233	3	US-10-103-327-21	Sequence 21, Appl
30	1212	96.7	1215	3	US-09-993-059-19	Sequence 19, Appl
31	1212	96.7	1215	3	US-10-103-327-19	Sequence 19, Appl
32	1176.4	93.8	1214	3	US-09-023-655-1001	Sequence 1001, Ap
33	1164	92.8	1164	3	US-09-626-127-13	Sequence 13, Appl
34	1164	92.8	1164	3	US-09-993-059-33	Sequence 33, Appl
35	1164	92.8	1164	3	US-10-103-327-33	Sequence 33, Appl
36	1159.4	92.5	1161	3	US-09-176-666-52	Sequence 52, Appl
37	1159.4	92.5	1164	3	US-09-176-666-51	Sequence 51, Appl
38	1159.4	92.5	1167	3	US-09-176-666-50	Sequence 50, Appl
39	1159.4	92.5	1170	3	US-09-176-666-49	Sequence 49, Appl
40	1159.4	92.5	1173	3	US-09-176-666-48	Sequence 48, Appl
41	1159.4	92.5	1176	3	US-09-176-666-47	Sequence 47, Appl
42	1159.4	92.5	1179	3	US-09-176-666-46	Sequence 46, Appl
43	1159.4	92.5	1182	3	US-09-176-666-45	Sequence 45, Appl
44	1159.4	92.5	1188	3	US-09-176-666-44	Sequence 44, Appl
45	1159.4	92.5	1194	3	US-09-176-666-43	Sequence 43, Appl

ALIGNMENTS

RESULT 1
US-09-993-059-15
; Sequence 15, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993, 059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-09-993-059-15

Query Match	100.0%;	Score 1254;	DB 3;	Length 1254;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1254;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	ATGCAGCTGAGGAAACCCAGAACCTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTCGGCC	60	
Db	1	ATGCAGCTGAGGAAACCCAGAACCTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTCGGCC	60	
QY	61	CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCCT	120	
Db	61	CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCCT	120	
QY	121	ACCATGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA	180	
Db	121	ACCATGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA	180	
QY	181	GATTCTGCATCAGTGAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC	240	
Db	181	GATTCTGCATCAGTGAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC	240	
QY	241	TGGAAGGATCGAGTTTATGAGTACCTCTGCTTGGATGGCTGCCCAAGA	300	
Db	241	TGGAAGGATCGAGTTTATGAGTACCTCTGCTTGGATGGCTGCCCAAGA	300	

```
301 GATTGAAAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db |
301 GATTGAAAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy |
361 GCTAATTATGTTACAGCAAGAGGAGCTGAAGCTAGGAGATTATGACAGATGTTGGAATAAA 420
Db |
361 GCTAATTATGTTACAGCAAGAGGAGCTGAAGCTAGGAGATTATGACAGATGTTGGAATAAA 420
Qy |
421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACACGACATGATGCCAGACCTTTGCT 480
Db |
421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACACGACATGATGCCAGACCTTTGCT 480
Qy |
481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGACAGATTTGGAATAATTG 540
Db |
481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGACAGATTTGGAATAATTG 540
Qy |
541 GCAGATGGTTATAAGCACATGCTTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 600
Db |
541 GCAGATGGTTATAAGCACATGCTTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 600
Qy |
601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db |
601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy |
661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATTTGATGATTTCTGGAAAGTATAAAG 720
Db |
661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATTTGATGATTTCTGGAAAGTATAAAG 720
Qy |
721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTCATGTTGTCGACACAGG 780
Db |
721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTCATGTTGTCGACACAGG 780
Qy |
781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
Db |
781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
Qy |
841 GTAACTCAGATGGCCCTTGGGCTATCATGGCTGCTCTTTTATGATCTAATGACCTC 900
Db |
841 GTAACTCAGATGGCCCTTGGGCTATCATGGCTGCTCTTTTATGATCTAATGACCTC 900
Qy |
901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGATAAGGACGTAATTTGCCATCAAT 960
Db |
901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGATAAGGACGTAATTTGCCATCAAT 960
Qy |
961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACTTTGAAGTGTG 1020
Db |
961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACTTTGAAGTGTG 1020
Qy |
1021 GAACGACCTCTCAGGCTTAGGCTGGCTGTAGCTATGATAAACCGGAGGAGATTGGT 1080
Db |
1021 GAACGACCTCTCAGGCTTAGGCTGGCTGTAGCTATGATAAACCGGAGGAGATTGGT 1080
Qy |
1081 GGACCTCGCTTATACCATCGAGTTGCTTCCCTGGGTAAGAGTGGCTCTGTAATCCT 1140
Db |
1081 GGACCTCGCTTATACCATCGAGTTGCTTCCCTGGGTAAGAGTGGCTCTGTAATCCT 1140
Qy |
1141 GCCTGCTTCAACACAGCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGATGAGCT 1200
Db |
1141 GCCTGCTTCAACACAGCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGATGAGCT 1200
Qy |
1201 TCAAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254
Db |
1201 TCAAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254
```

RESULT 2

```
US-10-103-327-15
; Sequence 15, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
```

```
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-10-103-327-15
```

```
Query Match 100.0%; Score 1254; DB 3; Length 1254;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTCGCTTCCTTCCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTCGCTTCCTTCCTGGCC 60
Qy 61 CTCGTTTCTTGGGACATCCCTGGGGCTAGAGCACTGGCAATGGATGGCAAGACGCT 120
Db 61 CTCGTTTCTTGGGACATCCCTGGGGCTAGAGCACTGGCAATGGATGGCAAGACGCT 120
Qy 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCC 180
Db 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCC 180
Qy 181 GATTCTCTGATCAGTGAAGAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Db 181 GATTCTCTGATCAGTGAAGAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Qy 241 TGAAGGATGCGAGTTATGAGTACCTCTGATGATGAGTGTGGATGGCTCCCAAGA 300
Db 241 TGAAGGATGCGAGTTATGAGTACCTCTGATGATGAGTGTGGATGGCTCCCAAGA 300
Qy 301 GATTTCAGAGGAGAGCTTCAGGCGAGACCTTCAGCGCTTTCCTCATGGATTCGCCAGCT 360
Db 301 GATTTCAGAGGAGAGCTTCAGGCGAGACCTTCAGGCGAGCTTTCCTCATGGATTCGCCAGCT 360
Qy 361 GCTAATTATGTTTACAGCAAGAGCTGAAGCTAGGAGTTTATGACAGATGTTGGAATAAA 420
Db 361 GCTAATTATGTTTACAGCAAGAGCTGAAGCTAGGAGTTTATGACAGATGTTGGAATAAA 420
Qy 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATGATGCCAGACCTTTGCT 480
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGAGCTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTG 540
Db 481 GACTGGGAGCTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTG 540
Qy 541 GCAGATGGTTATAAGCACATGCTCCTGGCCCTGAAATAGGACTGGCAGAAAGCATTTGTGTAC 600
Db 541 GCAGATGGTTATAAGCACATGCTCCTGGCCCTGAAATAGGACTGGCAGAAAGCATTTGTGTAC 600
Qy 601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATTTGATGATTTCTGGAAAGTATAAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATTTGATGATTTCTGGAAAGTATAAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTCATGTTGTCGACACGAGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTCATGTTGTCGACACGAGG 780
```


Db 721 AGTATCTTGACCTGGACATCTTTAAACGAGAGAAATGTTGATGTTGTCGACCAAGG 780
QY 781 GGTGGAAATGACCCAGATATGTTAGTGATTTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTGATTTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCAAGCTCTCCTTCAGGATAAGCAGCTAAATTTGCCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCAAGCTCTCCTTCAGGATAAGCAGCTAAATTTGCCATCAAT 960
QY 961 CAGGACCCCTTGGCAAGCAAGGTTACCACTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGCAAGCAAGGTTACCACTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 1021 GAACGACCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCGGACAGGATTTGGT 1080
Db 1021 GAACGACCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCGGACAGGATTTGGT 1080
QY 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTCTGTAATCCT 1140
Db 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTCTGTAATCCT 1140
QY 1141 GCCTGCTTCAATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCAATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTGTCTCAGCTATATA 1254
Db 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTGTCTCAGCTATATA 1254

RESULT 3

US-09-993-059-11
; Sequence 11, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-09-993-059-11

Query Match 99.9%; Score 1252.4; DB 3; Length 1266;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTTGGCC 60
Db 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTTGGCC 60
QY 61 CTGCTTCTTGGGACATCTCCTGGGCTTAGAGCTAGCAATGGATTGGCAAGAGCGCT 120
Db 61 CTGCTTCTTGGGACATCTCCTGGGCTTAGAGCTAGCAATGGATTGGCAAGAGCGCT 120
QY 121 ACATGGCTGGCTGACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCCA 180
Db 121 ACATGGCTGGCTGACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCCA 180

Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCCA 180
QY 181 GATTTCCTGCATCAGTGAGAGCTCTTTATGAGAGATGGCAGAGCTCATGTGCTCAGAAAGC 240
Db 181 GATTTCCTGCATCAGTGAGAGCTCTTTATGAGAGATGGCAGAGCTCATGTGCTCAGAAAGC 240
QY 241 TGAAGAGATGACAGGTTATGAGTACCTCTGCATTTGATGACTGTTGGATGGCTCCCAAGA 300
Db 241 TGAAGAGATGACAGGTTATGAGTACCTCTGCATTTGATGACTGTTGGATGGCTCCCAAGA 300
QY 301 GATTTCAGAGGAGAGACTTCAAGGAGAGCCCTCAGCGCTTTCCTCATGCGATTTGCGCAGCTA 360
Db 301 GATTTCAGAGGAGAGACTTCAAGGAGAGCCCTCAGCGCTTTCCTCATGCGATTTGCGCAGCTA 360
QY 361 GCTAATTTATGTTTCAACAGCAAGGACTGAAGCTAGGGATTTATGACAGATGTTGGAATAA 420
Db 361 GCTAATTTATGTTTCAACAGCAAGGACTGAAGCTAGGGATTTATGACAGATGTTGGAATAA 420
QY 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTTGATGATGATGCCAGACCTTTGCT 480
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTTGATGATGATGCCAGACCTTTGCT 480
QY 481 GACTGGGAGTACATCTGCTAAATTTTGAATTTGATGTTGTTACTGTGACAGTTTGGAAAAATTG 540
Db 481 GACTGGGAGTACATCTGCTAAATTTTGAATTTGATGTTGTTACTGTGACAGTTTGGAAAAATTG 540
QY 541 GCAGATGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTGTAC 600
Db 541 GCAGATGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
Db 601 TCCTGTGAGTGGCTCTTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
QY 661 CAGTACTGCAATCAGTGGGAGAAATTTTGTGACATTTGATGATGATGATGATGATGATGATGATGAT 720
Db 661 CAGTACTGCAATCAGTGGGAGAAATTTTGTGACATTTGATGATGATGATGATGATGATGATGATGAT 720
QY 721 AGTATCTTGGACTGGACATCTTTTAAACGAGAGAGAAATTTGTTGATGTTGTCGACCAAGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACGAGAGAGAAATTTGTTGATGTTGTCGACCAAGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTATTTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAATGACCCAGATATGTTAGTATTTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCAAGCTCTCCTTCAGGATAAGCAGCTAAATTTGCCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCAAGCTCTCCTTCAGGATAAGCAGCTAAATTTGCCATCAAT 960
QY 961 CAGGACCCCTTGGCAAGCAAGGTTACCACTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGCAAGCAAGGTTACCACTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 1021 GAACGACCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCGGACAGGATTTGGT 1080
Db 1021 GAACGACCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCGGACAGGATTTGGT 1080
QY 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTCTGTAATCCT 1140
Db 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTCTGTAATCCT 1140
QY 1141 GCCTGCTTCAATCACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCAATCACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTGTCTCAGCTATATA 1254
Db 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTGTCTCAGCTATATA 1254

```

RESULT 4
US-10-103-327-11
; Sequence 11, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-10-103-327-11

Query Match          99.9%; Score 1252.4; DB 3; Length 1266;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTCGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTCGGCC 60
QY 61 CTGCTTCTCGGACATCCCTGGGCTAGAGCACTGCAACATGATTTGGCAAGAGCGCT 120
DB 61 CTGCTTCTCGGACATCCCTGGGCTAGAGCACTGCAACATGATTTGGCAAGAGCGCT 120
QY 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
DB 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
QY 181 GATTCCTGCATAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
DB 181 GATTCCTGCATAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
QY 241 TGGAAAGGATGCAGGTTATGATGACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
DB 241 TGGAAAGGATGCAGGTTATGATGACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
QY 301 GATTCAGAAGCAGACTTCAGGCGAGCCCTCAGCGCTTTCCCTCATGGGATTCGCCAGCTA 360
DB 301 GATTCAGAAGCAGACTTCAGGCGAGCCCTCAGCGCTTTCCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAAATATGTTACAGCAAAAGGACTGAAGCTAGGATTTATGAGATTTGGAAATAAA 420
DB 361 GCTAAATATGTTACAGCAAAAGGACTGAAGCTAGGATTTATGAGATTTGGAAATAAA 420
QY 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACCTACGACATTCATGCCAGACTTTGCT 480
DB 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACCTACGACATTCATGCCAGACTTTGCT 480
QY 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTTACTGTGACAGTTTGGAAAATTTG 540
DB 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTTACTGTGACAGTTTGGAAAATTTG 540
QY 541 GCAGATGGTTATAGCAATGCTTCCTGGCCCTGAAATAGGACTGGCAGAACATTTGTATC 600
DB 541 GCAGATGGTTATAGCAATGCTTCCTGGCCCTGAAATAGGACTGGCAGAACATTTGTATC 600
QY 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660

US-09-993-059-7
; Sequence 7, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-09-993-059-7

Query Match          99.9%; Score 1252.4; DB 3; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTCGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTCGGCC 60
QY 61 CTGCTTCTCGGACATCCCTGGGCTAGAGCACTGCAACATGATTTGGCAAGAGCGCT 120
DB 61 CTGCTTCTCGGACATCCCTGGGCTAGAGCACTGCAACATGATTTGGCAAGAGCGCT 120
QY 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
DB 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
QY 181 GATTCCTGCATAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
DB 181 GATTCCTGCATAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
QY 241 TGGAAAGGATGCAGGTTATGATGACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
DB 241 TGGAAAGGATGCAGGTTATGATGACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
QY 301 GATTCAGAAGCAGACTTCAGGCGAGCCCTCAGCGCTTTCCCTCATGGGATTCGCCAGCTA 360
DB 301 GATTCAGAAGCAGACTTCAGGCGAGCCCTCAGCGCTTTCCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAAATATGTTACAGCAAAAGGACTGAAGCTAGGATTTATGAGATTTGGAAATAAA 420
DB 361 GCTAAATATGTTACAGCAAAAGGACTGAAGCTAGGATTTATGAGATTTGGAAATAAA 420
QY 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACCTACGACATTCATGCCAGACTTTGCT 480
DB 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACCTACGACATTCATGCCAGACTTTGCT 480
QY 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTTACTGTGACAGTTTGGAAAATTTG 540
DB 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTTACTGTGACAGTTTGGAAAATTTG 540
QY 541 GCAGATGGTTATAGCAATGCTTCCTGGCCCTGAAATAGGACTGGCAGAACATTTGTATC 600
DB 541 GCAGATGGTTATAGCAATGCTTCCTGGCCCTGAAATAGGACTGGCAGAACATTTGTATC 600
QY 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
```

Db 1 ATGCAGCTGAGAACCCAGAACATACATCTGGGCTGCGCGCTTGGCTTCCCTGCGC 60
Qy 61 CTGCTTTCCCTGGGACATCCCTGGGCTAGAGCACTGGAACAATGATGGCAAGAGCGCT 120
Db 61 CTGCTTTCCCTGGGACATCCCTGGGCTAGAGCACTGGAACAATGATGGCAAGAGCGCT 120
Qy 121 ACCATGGCTGGCTGCACTGGGAGCGCTCATGTGCAACCTTGACCTGCCAGGAGGCCA 180
Db 121 ACCATGGCTGGCTGCACTGGGAGCGCTCATGTGCAACCTTGACCTGCCAGGAGGCCA 180
Qy 181 GATTCTGTCATCAGTGAGAGCTCTTCATGAGATGCGAGCTCATGTGCTCAGAGGC 240
Db 181 GATTCTGTCATCAGTGAGAGCTCTTCATGAGATGCGAGCTCATGTGCTCAGAGGC 240
Qy 241 TGAAGGATCAGGTTATAGTAGTACCTCTGCAITGATGACTGTTGGATGGCTCCCAAGA 300
Db 241 TGAAGGATCAGGTTATAGTAGTACCTCTGCAITGATGACTGTTGGATGGCTCCCAAGA 300
Qy 301 GATTGAGAGGAGACTTTCAGGAGAGCCCTCAGCGCTTCCCTCATGGATTCGCCAGCTA 360
Db 301 GATTGAGAGGAGACTTTCAGGAGAGCCCTCAGCGCTTCCCTCATGGATTCGCCAGCTA 360
Qy 361 GCTAATTATGTTTACAGCAAGGACTGAAGCTAGGAGTTTATGCAGATGTTGGAATAAA 420
Db 361 GCTAATTATGTTTACAGCAAGGACTGAAGCTAGGAGTTTATGCAGATGTTGGAATAAA 420
Qy 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATACCTACGACATGATGCCAGAGCTTTGCT 480
Db 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATACCTACGACATGATGCCAGAGCTTTGCT 480
Qy 481 GACTGGGAGTAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Db 481 GACTGGGAGTAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
Qy 541 GCAGATGTTTATAGACATGCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Db 541 GCAGATGTTTATAGACATGCTTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Qy 601 TCCTGTGAGTGGCTCTTTATATGCTGCTGCTTCAAAGCCCAATTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATGCTGCTTCAAAGCCCAATTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGAAATTTGCTGCAATGATGATTCCTGGAAAGTATAAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTGCTGCAATGATGATTCCTGGAAAGTATAAAG 720
Qy 721 AGTATCTTGACATGCTTTTACCAGAGAGAAATGTTGATGTTGCTGACCAAGG 780
Db 721 AGTATCTTGACATGCTTTTACCAGAGAGAAATGTTGATGTTGCTGACCAAGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTATGCTGCTTCAAAGCCCAATTATACAGAAATCCGA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGCTTCAAAGCCCAATTATACAGAAATCCGA 840
Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTATGCTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTATGCTAATGACCTC 900
Qy 901 CGACATCAGCCCTCAGCCAAAGCTCTCTTCAGGATGAGGAGCTTAATGGCATCAAT 960
Db 901 CGACATCAGCCCTCAGCCAAAGCTCTCTTCAGGATGAGGAGCTTAATGGCATCAAT 960
Qy 961 CAGGACCCCTTGGGAGCAAGGCTTACAGCTTAGCAGGAGCAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGAGCAAGGCTTACAGCTTAGCAGGAGCAACTTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
Qy 1081 GGAACCTGCTCTTATACCATGCGAGTGTCTTCCCTGGGTTAAGAGTGGCTGCTAATCCT 1140

Db 1081 GGACCTGCTCTTATACCATGCGAGTGTCTTCCCTGGGTTAAGAGTGGCTGTAATCCT 1140
Qy 1141 GCCTGCTTCAATCACAGAGCTTCCCTGTGAAAGGAAGCTAGGGTCTTATGAACT 1200
Db 1141 GCCTGCTTCAATCACAGAGCTTCCCTGTGAAAGGAAGCTAGGGTCTTATGAACT 1200
Qy 1201 TCAAGGTTAAGAGTCAATTAATCCACAGGCACTGTTTGGCTTCAAGCTATAA 1254
Db 1201 TCAAGGTTAAGAGTCAATTAATCCACAGGCACTGTTTGGCTTCAAGCTAGAA 1254

RESULT 6

US-10-103-327-7
; Sequence 7, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-10-103-327-7

Query Match 99.9%; Score 1252.4; DB 3; Length 1278;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCAGCTGAGAACCCAGAACATACATCTGGGCTGCGCGCTTGGCTTCCCTGCGC 60

Db 1 ATGCAGCTGAGAACCCAGAACATACATCTGGGCTGCGCGCTTGGCTTCCCTGCGC 60

Qy 61 CTGCTTTCCCTGGGACATCCCTGGGCTAGAGCACTGGAACAATGATGGCAAGAGCGCT 120

Db 61 CTGCTTTCCCTGGGACATCCCTGGGCTAGAGCACTGGAACAATGATGGCAAGAGCGCT 120

Qy 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCAATGTGCAACCTTGACCTGCCAGGAGGCCA 180

Db 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCAATGTGCAACCTTGACCTGCCAGGAGGCCA 180

Qy 181 GATTCTGTCATCAGTGAGAGCTTTCATGGAGATGCGAGCTCATGTGCTCAGAGGC 240

Db 181 GATTCTGTCATCAGTGAGAGCTTTCATGGAGATGCGAGCTCATGTGCTCAGAGGC 240

Qy 241 TGAAGGATGCGAGTTATGATGACTCTGCAITGATGACTGTTGGATGGCTCCCAAGA 300

Db 241 TGAAGGATGCGAGTTATGATGACTCTGCAITGATGACTGTTGGATGGCTCCCAAGA 300

Qy 301 GATTTCAGAGGAGACTTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGATTCGCAAGCTA 360

Db 301 GATTTCAGAGGAGACTTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGATTCGCAAGCTA 360

Qy 361 GCTAATTATGTTTACAGCAAGGACTGAAGCTAGGAGTTTATGCAGATGTTGGAATAAA 420

Db 361 GCTAATTATGTTTACAGCAAGGACTGAAGCTAGGAGTTTATGCAGATGTTGGAATAAA 420

Qy 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACCTAGCAATGATGCCAGAGCTTTGCT 480

Db 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACCTAGCAATGATGCCAGAGCTTTGCT 480


```
Db 961 CAGGACCCCTTGGCAAGCAAGGTACAGCTTAGACAGGAGACAACCTTTGAAGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATATAACCGCAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATATAACCGCAGGAGATTGGT 1080
Qy 1081 GGACCTCGCTCTTATACATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Db 1081 GGACCTCGCTCTTATACATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Qy 1141 GCCTGCTTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTCTATGAATGACT 1200
Db 1141 GCCTGCTTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTCTATGAATGACT 1200
Qy 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGGCTTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGGCTTCAGCTAGAA 1254

RESULT 8
US-10-103-327-13
; Sequence 13, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1284)
US-10-103-327-13

Query Match 99.98; Score 1252.4; DB 3; Length 1284;
Best Local Similarity 99.98; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGCGCTTCGCTTCCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGCGCTTCGCTTCCTGGCC 60
Qy 61 CTGCTTTCTGGGACATCCCTGGGCTAGAGCACTGACCAATGATGGCAAGACGCT 120
Db 61 CTGCTTTCTGGGACATCCCTGGGCTAGAGCACTGACCAATGATGGCAAGACGCT 120
Qy 121 ACCATGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Db 121 ACCATGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Qy 181 GATTCTTCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 181 GATTCTTCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 241 TGAAGAGTCAAGTTATGATGATACCTCTGATGATGATGATGATGATGATGATGATGATGAT 300
Db 241 TGAAGAGTCAAGTTATGATGATACCTCTGATGATGATGATGATGATGATGATGATGATGAT 300
Qy 301 GATTTCAGAGGCACTTCAGGAGACCTCAGGCGCTTCTCATGGATTCGCGAGCTA 360
Db 301 GATTTCAGAGGCACTTCAGGAGACCTCAGGCGCTTCTCATGGATTCGCGAGCTA 360
```

```
Qy 361 GCTAATTTATGTTTCAAGCAAGGACGCTGAAGCTAGGGATTTATGAGATGTTGGAATAAA 420
Db 361 GCTAATTTATGTTTCAAGCAAGGACGCTGAAGCTAGGGATTTATGAGATGTTGGAATAAA 420
Qy 421 ACTTGGCGAGCTTCCCTGGAGCTTTGGATACCTAGCAGATGATGATGATGATGATGATGAT 480
Db 421 ACTTGGCGAGCTTCCCTGGAGCTTTGGATACCTAGCAGATGATGATGATGATGATGATGAT 480
Qy 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTGTGTTACTGTGACAGTTCGAAAAATTTG 540
Db 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTGTGTTACTGTGACAGTTCGAAAAATTTG 540
Qy 541 GCAGATGGTTATTAAGCAATGCTTGGCCCTGAAATAGGACTGGCAGAGCAATTTGTGTAC 600
Db 541 GCAGATGGTTATTAAGCAATGCTTGGCCCTGAAATAGGACTGGCAGAGCAATTTGTGTAC 600
Qy 601 TCCTGTGAGTGGCTCTTTATATATGAGGCTCTTTTAAAGCCCAATTAACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATATGAGGCTCTTTTAAAGCCCAATTAACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGAAAATTTTGTGCTGACATTTGATGATTTCTCGAAAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAAATTTTGTGCTGACATTTGATGATTTCTCGAAAAAGTATAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGATTTGCTGGACGAGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGATTTGCTGGACGAGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTATGAGTGGCACTTTGGCCCTCAGCTCGAATCAGCA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGAGTGGCACTTTGGCCCTCAGCTCGAATCAGCA 840
Qy 841 GTAACTCAGATGCGCTCTGGGCTATCATGGCTGCTCTTTTATTTATGATGATGATGATGAT 900
Db 841 GTAACTCAGATGCGCTCTGGGCTATCATGGCTGCTCTTTTATTTATGATGATGATGATGAT 900
Qy 901 CGACATCATGAGCCCTCAAGCAAGCTCTCTTCAAGGATAAGGACGTAATTTGCCATCAAT 960
Db 901 CGACATCATGAGCCCTCAAGCAAGCTCTCTTCAAGGATAAGGACGTAATTTGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGTACCACTTAGACAGGAGCACTTTGAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTACCACTTAGACAGGAGCACTTTGAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATATAACCGCAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATATAACCGCAGGAGATTGGT 1080
Qy 1081 GGACCTCGCTCTTATACATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Db 1081 GGACCTCGCTCTTATACATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Qy 1141 GCCTGCTTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTCTATGAATGACT 1200
Db 1141 GCCTGCTTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTCTATGAATGACT 1200
Qy 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGGCTTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGGCTTCAGCTAGAA 1254
```

RESULT 9

```
US-09-993-059-3
; Sequence 3, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
```

```
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; LOCATION: (1)...(1290)
US-09-993-059-3

Query Match          99.9%; Score 1252.4; DB 3; Length 1290;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCGCTTCCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCGCTTCCTGGCC 60
QY 61 CTGCTTTCTCTGGGACATCCCTGGGCTTAGAGCACTGGACAATGGATGGCAAGGACGCT 120
DB 61 CTGCTTTCTCTGGGACATCCCTGGGCTTAGAGCACTGGACAATGGATGGCAAGGACGCT 120
QY 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCCA 180
DB 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCCA 180
QY 181 GATTCTGTCATCAGTGAAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
DB 181 GATTCTGTCATCAGTGAAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 241 TGAAGGATGACAGTTATGATGATACCTCTGCATTGATGACTGTTGGATGGCTCCCAAGA 300
DB 241 TGAAGGATGACAGTTATGATGATACCTCTGCATTGATGACTGTTGGATGGCTCCCAAGA 300
QY 301 GATTCAAGGAGACTTCAGGAGACCCCTCAGGCTTTCCTCATGGATTCGCCAGCTA 360
DB 301 GATTCAAGGAGACTTCAGGAGACCCCTCAGGCTTTCCTCATGGATTCGCCAGCTA 360
QY 361 GCTAATTTATCTCACAGCAAGAGCTGAAGCTAGGAGATTTATGAGATGTTGGAATAAA 420
DB 361 GCTAATTTATCTCACAGCAAGAGCTGAAGCTAGGAGATTTATGAGATGTTGGAATAAA 420
QY 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTAGACATGATGATGCCAGACCTTTGCT 480
DB 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTAGACATGATGATGCCAGACCTTTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTTACTGTGACAGTTTGGAAAATTG 540
DB 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTTACTGTGACAGTTTGGAAAATTG 540
QY 541 GCAGATGTTTATAGACATGCTCTGGCCCTGAAATAGGACTGSCAGAACCTTGTGTAC 600
DB 541 GCAGATGTTTATAGACATGCTCTGGCCCTGAAATAGGACTGSCAGAACCTTGTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGGCCAATTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGGCCAATTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTGGAAAAGTATAAG 720
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTGGAAAAGTATAAG 720
QY 721 AGTATCTTGACTGGACATCTTTTAAACAGAGAGAAATTTGATGTTGCTGGACACAGG 780
DB 721 AGTATCTTGACTGGACATCTTTTAAACAGAGAGAAATTTGATGTTGCTGGACACAGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
DB 781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
```

```
QY 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTAATGACCTC 900
DB 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATGGCATCAAT 960
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATGGCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCGCAGGAGATTGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCGCAGGAGATTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTGGGTAAAGGAGTGGCTGTAACTCT 1140
DB 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTGGGTAAAGGAGTGGCTGTAACTCT 1140
QY 1141 GCCTGCTTCATCACACAGCTCCCTCTGTGAAAGGAAGCTAGGTTCTATGAATGGACT 1200
DB 1141 GCCTGCTTCATCACACAGCTCCCTCTGTGAAAGGAAGCTAGGTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTGTCTTCAGCTATAA 1254
DB 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTGTCTTCAGCTATAA 1254

RESULT 10
US-10-103-327-3
; Sequence 3, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCES: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; LOCATION: (1)...(1290)
US-10-103-327-3

Query Match          99.9%; Score 1252.4; DB 3; Length 1290;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCGCTTCCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCGCTTCCTGGCC 60
QY 61 CTGCTTTCTCTGGGACATCCCTGGGCTTAGAGCACTGGACAATGGATGGCAAGGACGCT 120
DB 61 CTGCTTTCTCTGGGACATCCCTGGGCTTAGAGCACTGGACAATGGATGGCAAGGACGCT 120
QY 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCCA 180
DB 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCCA 180
QY 181 GATTCTGTCATCAGTGAAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
DB 181 GATTCTGTCATCAGTGAAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 241 TGAAGGATGACAGTTATGATGATACCTCTGCATTGATGACTGTTGGATGGCTCCCAAGA 300
DB 241 TGAAGGATGACAGTTATGATGATACCTCTGCATTGATGACTGTTGGATGGCTCCCAAGA 300
QY 301 GATTCAAGGAGACTTCAGGAGACCCCTCAGGCTTTCCTCATGGATTCGCCAGCTA 360
DB 301 GATTCAAGGAGACTTCAGGAGACCCCTCAGGCTTTCCTCATGGATTCGCCAGCTA 360
QY 361 GCTAATTTATCTCACAGCAAGAGCTGAAGCTAGGAGATTTATGAGATGTTGGAATAAA 420
DB 361 GCTAATTTATCTCACAGCAAGAGCTGAAGCTAGGAGATTTATGAGATGTTGGAATAAA 420
QY 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTAGACATGATGATGCCAGACCTTTGCT 480
DB 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTAGACATGATGATGCCAGACCTTTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTTACTGTGACAGTTTGGAAAATTG 540
DB 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTTACTGTGACAGTTTGGAAAATTG 540
QY 541 GCAGATGTTTATAGACATGCTCTGGCCCTGAAATAGGACTGSCAGAACCTTGTGTAC 600
DB 541 GCAGATGTTTATAGACATGCTCTGGCCCTGAAATAGGACTGSCAGAACCTTGTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGGCCAATTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGGCCAATTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTGGAAAAGTATAAG 720
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTGGAAAAGTATAAG 720
QY 721 AGTATCTTGACTGGACATCTTTTAAACAGAGAGAAATTTGATGTTGCTGGACACAGG 780
DB 721 AGTATCTTGACTGGACATCTTTTAAACAGAGAGAAATTTGATGTTGCTGGACACAGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
DB 781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
```

```
Db 181 GATTCTCTGATCAGTGAGAGCTCTTTCATGGAGATGCGAGAGCTCATGGTCTCAGAAGGC 240
Qy 241 TGGAGAGATGACAGTTATAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAAGA 300
Db 241 TGGAGAGATGACAGTTATAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAAGA 300
Qy 301 GATTCAAGAGCAGACTTCAGGACAGCCCTCAGCGCTTCTCATGGGATTCGCAGCTA 360
Db 301 GATTCAAGAGCAGACTTCAGGACAGCCCTCAGCGCTTCTCATGGGATTCGCAGCTA 360
Qy 361 GCTAATTTATGTTTCACAGCAAGGACTGAAGCTAGGGATTTATGACAGATGTTGGAATAAA 420
Db 361 GCTAATTTATGTTTCACAGCAAGGACTGAAGCTAGGGATTTATGACAGATGTTGGAATAAA 420
Qy 421 ACCTGCGCAGGCTTCCCTGGAGATTTGGATPACTACGACATGATGATGCCAGACCTTTGCT 480
Db 421 ACCTGCGCAGGCTTCCCTGGAGATTTGGATPACTACGACATGATGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAATTTG 540
Db 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAATTTG 540
Qy 541 GCAGATGGTTTATAGACATGCTCTGGCCCTGAAATAGGACTGGCAGAAAGCATTTGTGTAC 600
Db 541 GCAGATGGTTTATAGACATGCTCTGGCCCTGAAATAGGACTGGCAGAAAGCATTTGTGTAC 600
Qy 601 TCCTGTGAGTGGCTCTTTTATATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTTATATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGCAATTTTGTGACATTTGATGATTTCTCTGGAAAAGTATAAAG 720
Db 661 CAGTACTGCAATCACTGGCGCAATTTTGTGACATTTGATGATTTCTCTGGAAAAGTATAAAG 720
Qy 721 AGTATCTTGACATGGACATCTTTTAAACAGAGAGAAATTTGTGATGTTGTCGACACAGGG 780
Db 721 AGTATCTTGACATGGACATCTTTTAAACAGAGAGAAATTTGTGATGTTGTCGACACAGGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTATGTCGAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGTCGAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATGCTATGATGACCTC 900
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATGCTATGATGACCTC 900
Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGGTACACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGGCTGGGCTGATGCTATGATATAACCGGAGAGATGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGGCTGGGCTGATGCTATGATATAACCGGAGAGATGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCTCTGGTAAAGAGTGGCTCTGTAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCTCTGGTAAAGAGTGGCTCTGTAATCCT 1140
Qy 1141 GCCTGCTTCATACACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGAAT 1200
Db 1141 GCCTGCTTCATACACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGAAT 1200
Qy 1201 TCAAGGTTAAGAAGTACATAAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAAGTACATAAATCCACAGGCACTGTTTGTCTCAGCTATAGAA 1254
```

RESULT 11

US-09-993-059-9

```
; Sequence 9, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1296
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1296)
US-09-993-059-9
```

```
Query Match 99.9%; Score 1252.4; DB 3; Length 1296;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCGCTTCCTGGCC 60
Db 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCGCTTCCTGGCC 60
Qy 61 CTGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTTGGCAGGACGCT 120
Db 61 CTGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTTGGCAGGACGCT 120
Qy 121 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTCCAGGAAGAGCCA 180
Db 121 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTCCAGGAAGAGCCA 180
Qy 181 GATTCTGCACTAGTGAGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
Db 181 GATTCTGCACTAGTGAGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
Qy 241 TGAAGGATGACAGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAAGA 300
Db 241 TGAAGGATGACAGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAAGA 300
Qy 301 GATTCAAGAGCAGACTTCAGGACAGCCCTCAGCGCTTCTCATGGGATTCGCAGCTA 360
Db 301 GATTCAAGAGCAGACTTCAGGACAGCCCTCAGCGCTTCTCATGGGATTCGCAGCTA 360
Qy 361 GCTAATTTATGTTTCACAGCAAGGACTGAAGCTAGGGATTTATGACAGATGTTGGAATAAA 420
Db 361 GCTAATTTATGTTTCACAGCAAGGACTGAAGCTAGGGATTTATGACAGATGTTGGAATAAA 420
Qy 421 ACCTGCGCAGGCTTCCCTGGAGATTTGGATPACTACGACATGATGATGCCAGACCTTTGCT 480
Db 421 ACCTGCGCAGGCTTCCCTGGAGATTTGGATPACTACGACATGATGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAATTTG 540
Db 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAATTTG 540
Qy 541 GCAGATGGTTTATAGACATGCTCTGGCCCTGAAATAGGACTGGCAGAAAGCATTTGTGTAC 600
Db 541 GCAGATGGTTTATAGACATGCTCTGGCCCTGAAATAGGACTGGCAGAAAGCATTTGTGTAC 600
Qy 601 TCCTGTGAGTGGCTCTTTTATATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTTATATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGCAATTTTGTGACATTTGATGATTTCTCTGGAAAAGTATAAAG 720
Db 661 CAGTACTGCAATCACTGGCGCAATTTTGTGACATTTGATGATTTCTCTGGAAAAGTATAAAG 720
```


QY	721	AGTATCTTGGACTGGACATCTTTTAAACGAGGAGAGAAATGTTGATGTTGCTCGACAGGG	780	Db	61	CTCGTTCTCTGGGACATCCTCGGGCTAGAGCACTGGCAATGGATTGGCAAGAGCGCT	120
Db	721	AGTATCTTGGACTGGACATCTTTTAAACGAGGAGAGAAATGTTGATGTTGCTCGACAGGG	780	QY	121	ACATGGGGCTGGCTGCATCGGAGCGCTTCATGTGCAACTTTCAGCTGCCAGGAAGAGCCA	180
QY	781	GTTTGGAAATGACCCAGATATGTTAGTATGTTGGCACTTTGGGCTCAGCTGGAAATCAGCAA	840	Db	121	ACCATGGGCTGGCTGCATCGGAGCGCTTCATGTGCAACTTTCAGCTGCCAGGAAGAGCCA	180
Db	781	GTTTGGAAATGACCCAGATATGTTAGTATGTTGGCACTTTGGGCTCAGCTGGAAATCAGCAA	840	QY	181	GATTCCTGCATCAGTGTAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAAGGC	240
QY	841	GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGTCTAATGACCTC	900	Db	181	GATTCCTGCATCAGTGTAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAAGGC	240
Db	841	GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGTCTAATGACCTC	900	QY	241	TGGAAGGATGCGAGTTTATCAGTACTCCTCTGCATTTGATGACTGTGTGGATGGCTCCCCAAAGA	300
QY	901	CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCGCATCAAT	960	Db	241	TGGAAGGATGCGAGTTTATCAGTACTCCTCTGCATTTGATGACTGTGTGGATGGCTCCCCAAAGA	300
Db	901	CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCGCATCAAT	960	QY	301	GATTCAGAAAGGAGACTTCAGGAGACCTTCAGGCGCTTTCCTCATGGATTCGCCAGCTA	360
QY	961	CAGGACCCCTTGGGCAAGCAAGGTACCACTTAGACAGGAGACAACCTTGAAGTGTGG	1020	Db	301	GATTCAGAAAGGAGACTTCAGGAGACCTTCAGGCGCTTTCCTCATGGATTCGCCAGCTA	360
Db	961	CAGGACCCCTTGGGCAAGCAAGGTACCACTTAGACAGGAGACAACCTTGAAGTGTGG	1020	QY	361	GCTAATTATGTTTACAGCAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAATATAA	420
QY	1021	GAAACCTCTCTCAGGCTTAGCGCTGTAGCTATGATATAAACCGCAGGAGATTGGT	1080	Db	361	GCTAATTATGTTTACAGCAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAATATAA	420
Db	1021	GAAACCTCTCTCAGGCTTAGCGCTGTAGCTATGATATAAACCGCAGGAGATTGGT	1080	QY	421	ACCTGCGCAGGCTTCCCTGGAGTTTGGATACTACGACATTCAGTATGCCAGACCTTTGCT	480
QY	1081	GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT	1140	Db	421	ACCTGCGCAGGCTTCCCTGGAGTTTGGATACTACGACATTCAGTATGCCAGACCTTTGCT	480
Db	1081	GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT	1140	QY	481	GACTGGGGAGTAGACTCTGCTTAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTG	540
QY	1141	GCTGCTTATCACAGCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGCACT	1200	Db	481	GACTGGGGAGTAGACTCTGCTTAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTG	540
Db	1141	GCTGCTTATCACAGCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGCACT	1200	QY	541	GCAGATGTTTATAAGCACATGCTCTTGGCCCTCAATAGGACTGGCAGAAACATTTGTGTC	600
QY	1201	TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTATATA	1254	Db	541	GCAGATGTTTATAAGCACATGCTCTTGGCCCTCAATAGGACTGGCAGAAACATTTGTGTC	600
Db	1201	TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTATATA	1254	QY	601	TCCTGTGAGTGGCTCTTTATATGTTGTCCTTCAAAAGCCCAATATATACAGAAATCCGA	660
RESULT 12							
US-10-103-327-9							
; Sequence 9, Application US/10103327							
; Patent No. 6890748							
; GENERAL INFORMATION:							
; APPLICANT: GARGER, Stephen A.							
; APPLICANT: TURPEN, Thomas H.							
; APPLICANT: KUMAGAI, Monto H.							
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN							
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION							
; FILE REFERENCE: 008010087CPUS06							
; CURRENT APPLICATION NUMBER: US/10/103,327							
; PRIOR FILING DATE: 2002-03-20							
; PRIOR APPLICATION NUMBER: US/09/993,059							
; PRIOR FILING DATE: 2001-11-13							
; NUMBER OF SEQ ID NOS: 37							
; SOFTWARE: FastSeq for Windows Version 4.0							
; SEQ ID NO 9							
; LENGTH: 1296							
; TYPE: DNA							
; ORGANISM: Homo sapiens							
; FEATURE:							
; NAME/KEY: CDS							
; LOCATION: (1)... (1296)							
US-10-103-327-9							
Query Match							
Best Local Similarity 99.9%; Score 1252.4; DB 3; Length 1296;							
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;							
QY	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC	60				
Db	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC	60				
QY	61	CTCGTTCTCGGACATCCCTGGGGCTAGAGCAATGGATTGGCAAGAGCGCT	120				

Query Match		99.9%;	Score 1252.4;	DB 3;	Length 1308;	
Best Local Similarity		99.9%;	Pred. No. 0;			
Matches 1253;		Conservative	0;	Mismatches	1;	Indels 0; Gaps 0;
Qy	1	ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTGGCTTCTCTGGCC	60			
Db	1	ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTGGCTTCTCTGGCC	60			
Qy	61	CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTTGGCAAGAGCGCT	120			
Db	61	CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTTGGCAAGAGCGCT	120			
Qy	121	ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACTTGGACTGCCAGGAAGGCCA	180			
Db	121	ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACTTGGACTGCCAGGAAGGCCA	180			
Qy	181	GATTCCTGTCATCAGTCAGAAAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC	240			
Db	181	GATTCCTGTCATCAGTCAGAAAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC	240			
Qy	241	TGGAAGGATGCGAGTTATGAGTACCTCTGCAATGATGACTTTGGATGGCTCCCAAGA	300			
Db	241	TGGAAGGATGCGAGTTATGAGTACCTCTGCAATGATGACTTTGGATGGCTCCCAAGA	300			
Qy	301	GATTCAGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGATTCGCCAGCTA	360			
Db	301	GATTCAGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGATTCGCCAGCTA	360			
Qy	361	GCTAATTATGTTTCCACAGCAAGGACTGMACTAGGATTTATGCAGATGTTGGAATAAA	420			
Db	361	GCTAATTATGTTTCCACAGCAAGGACTGMACTAGGATTTATGCAGATGTTGGAATAAA	420			
Qy	421	ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACACGACATTCGATCCAGACCTTTGCT	480			
Db	421	ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACACGACATTCGATCCAGACCTTTGCT	480			
Qy	481	GACTGGGGAGTAGACTCTGCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTG	540			
Db	481	GACTGGGGAGTAGACTCTGCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTG	540			
Qy	541	GCAGATGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTATC	600			
Db	541	GCAGATGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTATC	600			
Qy	601	TCTGTGAGTGGCTCTTTATATGTGGCCCTTTTCAAAGCCCAATTATACAGAAATCCGA	660			
Db	601	TCTGTGAGTGGCTCTTTATATGTGGCCCTTTTCAAAGCCCAATTATACAGAAATCCGA	660			
Qy	661	CAGTACTGCAATCACTGGCGAAATTTTGTGCAATTTGATGATTTCTTGGAAAGTATAAG	720			
Db	661	CAGTACTGCAATCACTGGCGAAATTTTGTGCAATTTGATGATTTCTTGGAAAGTATAAG	720			
Qy	721	AGTATCTTGAGCTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGTCGACACAGGG	780			
Db	721	AGTATCTTGAGCTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGTCGACACAGGG	780			
Qy	781	GGTTGGAATGACCCAGATATGTTAGTGAATTTGGCAACTTTGGCTTCAGCTGGAATACGAA	840			
Db	781	GGTTGGAATGACCCAGATATGTTAGTGAATTTGGCAACTTTGGCTTCAGCTGGAATACGAA	840			
Qy	841	GTAATCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTCATATGACCTC	900			
Db	841	GTAATCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTCATATGACCTC	900			
Qy	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT	960			
Db	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT	960			
Qy	961	CAGGACCCCTTGGCAAGAGGGTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG	1020			
Db	961	CAGGACCCCTTGGCAAGAGGGTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG	1020			
Qy	1021	GAAGGACCTCTTCAGGCTTAGCTGGGCTGATGATGATTAACCGGAGGAGATTGGT	1080			

Db	1021	GAAGGACCTCTTCAGGCTTAGCTGGGCTGATGATATATAAACCGGAGGAGATTGGT	1080			
Qy	1081	GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCCTGTAATCCT	1140			
Db	1081	GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCCTGTAATCCT	1140			
Qy	1141	GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT	1200			
Db	1141	GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT	1200			
Qy	1201	TCAAGGTTAAGAGTCACTATAATCCACAGGCACTGTTTGTCTCAGCTATATA	1254			
Db	1201	TCAAGGTTAAGAGTCACTATAATCCACAGGCACTGTTTGTCTCAGCTATATA	1254			
RESULT 15						
US-08-928-881-18						
; Sequence 18, Application US/08928881						
; Patent No. 6083725						
; GENERAL INFORMATION:						
; APPLICANT: Seiden et al., Richard F.						
; TITLE OF INVENTION: THERAPY FOR ALPHA-GALACTOSIDASE A						
; TITLE OF INVENTION: DEFICIENCY						
; NUMBER OF SEQUENCES: 28						
; CORRESPONDENCE ADDRESS:						
; ADDRESSEE: Fish & Richardson						
; STREET: 225 Franklin Street						
; CITY: Boston						
; STATE: MA						
; COUNTRY: USA						
; ZIP: 02110-2804						
; COMPUTER READABLE FORM: disk						
; MEDIUM TYPE: Floppy disk						
; COMPUTER: IBM PC compatible						
; OPERATING SYSTEM: PC-DOS/MS-DOS						
; SOFTWARE: Patent In Release #1.0, Version #1.30						
; CURRENT APPLICATION DATA:						
; APPLICATION NUMBER: US/08/928,881						
; FILING DATE:						
; CLASSIFICATION: 435						
; ATTORNEY/AGENT INFORMATION:						
; NAME: Fraser, Janis K.						
; REGISTRATION NUMBER: 34,819						
; REFERENCE/DOCKET NUMBER: 07236/003001						
; TELECOMMUNICATION INFORMATION:						
; TELEPHONE: 617/542-5070						
; TELEFAX: 617/542-8906						
; INFORMATION FOR SEQ ID NO: 18:						
; SEQUENCE CHARACTERISTICS:						
; LENGTH: 1343 base pairs						
; TYPE: nucleic acid						
; STRANDEDNESS: single						
; TOPOLOGY: linear						
US-08-928-881-18						
Query Match						
Best Local Similarity						
Matches 1253;						
99.9%; Score 1252.4; DB 3; Length 1343;						
Pred. No. 0;						
Mismatches 1;						
Indels 0; Gaps 0;						
Qy	1	ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTGGCTTCTCTGGCC	60			
Db	36	ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTGGCTTCTCTGGCC	95			
Qy	61	CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTTGGCAAGAGCGCT	120			
Db	96	CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTTGGCAAGAGCGCT	155			
Qy	121	ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGGACTGCCAGGAAGGCCA	180			
Db	156	ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGGACTGCCAGGAAGGCCA	215			
Qy	181	GATTCCTGTCATCAGTCAGAAAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC	240			

Job time : 266 secs

Db 216 GATTCCTGCATCAGTGAGAGCTCTCATGGAGATGCGAGCTCATGGTCTCAGAAGGC 275
Qy
Db 241 TGAAGAGATCAGGCTTATGAGTACTCTGTCATTGATGACTGTGTGGATGGCTCCCCAAAGA 300
Qy
Db 276 TGGAGGATCGAGGTATGAGTACCTCTGTCATTGATGACTGTGTGGATGGCTCCCCAAAGA 335
Qy
Db 301 GATTCAGAAGCGAGACTTCAGGCGAGACCCCTCAGCGCTTTCTCATNGGGATTCGCCAGCTA 360
Qy
Db 336 GATTCAGAAGCGAGACTTCAGGCGAGACCCCTCAGCGCTTTCTCATNGGGATTCGCCAGCTA 395
Qy
Db 361 GCTAAATTTATGTTACAGCAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAAATAAA 420
Qy
Db 396 GCTAAATTTATGTTACAGCAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAAATAAA 455
Qy
Db 421 ACCTGCCGAGGCTTCCCTGGGAGTTTGGGATACCTACGACATGATGCCAGACCTTTGCT 480
Qy
Db 456 ACCTGCCGAGGCTTCCCTGGGAGTTTGGGATACCTACGACATGATGCCAGACCTTTGCT 515
Qy
Db 481 GACTGGGGAGTAGACTGCTTAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAAATTG 540
Qy
Db 516 GACTGGGGAGTAGACTGCTTAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAAATTG 575
Qy
Db 541 GCAGATGGTTATAGCACATGCTCTGGCCCTGNAAGGACTGGCAGAGCATTTGTAC 600
Qy
Db 576 GCAGATGGTTATAGCACATGCTCTGGCCCTGNAAGGACTGGCAGAGCATTTGTAC 635
Qy
Db 601 TCCTGTGAGTGGCTCTTTATATGTTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 660
Qy
Db 636 TCCTGTGAGTGGCTCTTTATATGTTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 695
Qy
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAAATATAAAG 720
Qy
Db 696 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAAATATAAAG 755
Qy
Db 721 AGTATCTTGGACTGGACATCTTTTAAACCAGGAGAAATTTGATGATTCCTGGAAAAATATAAAG 780
Qy
Db 756 AGTATCTTGGACTGGACATCTTTTAAACCAGGAGAAATTTGATGATTCCTGGAAAAATATAAAG 815
Qy
Db 781 GGTGTGAATGACCCAGATATGTTAGTATGTTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy
Db 816 GGTGTGAATGACCCAGATATGTTAGTATGTTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 875
Qy
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATGCTAATGACCTC 900
Qy
Db 876 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATGCTAATGACCTC 935
Qy
Db 901 CGACACATCAGCCCTCAGCCCAAGCTCTCTTCAGGATAGGACGTAATTCGCATCAAT 960
Qy
Db 936 CGACACATCAGCCCTCAGCCCAAGCTCTCTTCAGGATAGGACGTAATTCGCATCAAT 995
Qy
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGCAGGAGACAACCTTTGAAGTGTGG 1020
Qy
Db 996 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGCAGGAGACAACCTTTGAAGTGTGG 1055
Qy
Db 1021 GAACGACCTCTCTCAGGCTTTAGCCTGGGCTGTAGCTATGATAAACCGCAGGAGATTGGT 1080
Qy
Db 1056 GAACGACCTCTCTCAGGCTTTAGCCTGGGCTGTAGCTATGATAAACCGCAGGAGATTGGT 1115
Qy
Db 1081 GGACCTGGCTCTTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTATCCT 1140
Qy
Db 1116 GGACCTGGCTCTTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTATCCT 1175
Qy
Db 1141 GCCTGTCTTATCAGACATCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGACT 1200
Qy
Db 1176 GCCTGTCTTATCAGACATCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGACT 1235
Qy
Db 1201 TCAAGGTTAAGAGTCAATAAATCCCAAGGACCTGTTTTGCTTCAGCTATAA 1254
Qy
Db 1236 TCAAGGTTAAGAGTCAATAAATCCCAAGGACCTGTTTTGCTTCAGCTAGAA 1289

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 31, 2005, 19:26:56 ; Search time 1160 Seconds
(without alignments)
8939.483 Million cell updates/sec

Title: US-10-602-220-15
Perfect score: 1254
Sequence: 1 atgcagctgagggaaccaga.....ctgttttgcagctataa 1254

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA_Main:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1254	100.0	1254	3	US-09-993-059-15
2	1254	100.0	1254	5	US-10-103-327-15
3	1254	100.0	1254	7	US-10-602-219-15
4	1254	100.0	1254	7	US-10-602-220-15
5	1254	100.0	1254	8	US-10-851-388-15
6	1254	100.0	1254	9	US-10-984-389-15
7	1252.4	99.9	1266	3	US-09-993-059-11
8	1252.4	99.9	1266	5	US-10-103-327-11
9	1252.4	99.9	1266	7	US-10-602-219-11
10	1252.4	99.9	1266	7	US-10-602-220-11
11	1252.4	99.9	1266	8	US-10-851-388-11
12	1252.4	99.9	1266	9	US-10-984-389-11
13	1252.4	99.9	1278	3	US-09-993-059-7
14	1252.4	99.9	1278	5	US-10-103-327-7
15	1252.4	99.9	1278	7	US-10-602-219-7
16	1252.4	99.9	1278	7	US-10-602-220-7
17	1252.4	99.9	1278	8	US-10-851-388-7
18	1252.4	99.9	1278	9	US-10-984-389-7
19	1252.4	99.9	1284	3	US-09-993-059-13
20	1252.4	99.9	1284	5	US-10-103-327-13
21	1252.4	99.9	1284	7	US-10-602-219-13
22	1252.4	99.9	1284	7	US-10-602-220-13
23	1252.4	99.9	1284	8	US-10-851-388-13

24	1252.4	99.9	1284	9	US-10-984-389-13	Sequence 13, Appl
25	1252.4	99.9	1290	3	US-09-993-059-3	Sequence 3, Appl
26	1252.4	99.9	1290	5	US-10-103-327-3	Sequence 3, Appl
27	1252.4	99.9	1290	7	US-10-602-219-3	Sequence 3, Appl
28	1252.4	99.9	1290	7	US-10-602-220-3	Sequence 3, Appl
29	1252.4	99.9	1290	7	US-10-411-037-67	Sequence 67, Appl
30	1252.4	99.9	1290	7	US-10-411-026-67	Sequence 67, Appl
31	1252.4	99.9	1290	7	US-10-410-962-67	Sequence 67, Appl
32	1252.4	99.9	1290	7	US-10-411-049-67	Sequence 67, Appl
33	1252.4	99.9	1290	7	US-10-410-930-67	Sequence 67, Appl
34	1252.4	99.9	1290	7	US-10-410-997-67	Sequence 67, Appl
35	1252.4	99.9	1290	7	US-10-411-012-67	Sequence 67, Appl
36	1252.4	99.9	1290	7	US-10-410-913-67	Sequence 67, Appl
37	1252.4	99.9	1290	8	US-10-851-388-3	Sequence 3, Appl
38	1252.4	99.9	1290	8	US-10-410-980-67	Sequence 67, Appl
39	1252.4	99.9	1290	9	US-10-410-987-67	Sequence 67, Appl
40	1252.4	99.9	1290	9	US-10-984-389-3	Sequence 3, Appl
41	1252.4	99.9	1296	3	US-09-993-059-9	Sequence 9, Appl
42	1252.4	99.9	1296	5	US-10-103-327-9	Sequence 9, Appl
43	1252.4	99.9	1296	7	US-10-602-219-9	Sequence 9, Appl
44	1252.4	99.9	1296	7	US-10-602-220-9	Sequence 9, Appl
45	1252.4	99.9	1296	8	US-10-851-388-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-09-993-059-15
; Sequence 15, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-09-993-059-15

Query Match	100.0%	Score 1254;	DB 3;	Length 1254;	
Best Local Similarity	100.0%	Pred. No. 0;			
Matches 1254;	Conservative	0;	Mismatches	0;	Gaps
QY	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGCCTTCCCTGGCC	60		
Db	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGCCTTCCCTGGCC	60		
QY	61	CTCGTTTCTGGGACATCTCTGGGCTAGAGCACTGGACATGGATGGCAAGACGCT	120		
Db	61	CTCGTTTCTGGGACATCTCTGGGCTAGAGCACTGGACATGGATGGCAAGACGCT	120		
QY	121	ACCATGGCTGGCTGACCTGGGAGCGCTTTCATGTGCAACCTTGACTGCCAGGAGAGCCCA	180		
Db	121	ACCATGGCTGGCTGACCTGGGAGCGCTTTCATGTGCAACCTTGACTGCCAGGAGAGCCCA	180		
QY	181	GATTCCTGCATCAGTGAGAGCTTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAGGC	240		
Db	181	GATTCCTGCATCAGTGAGAGCTTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAGGC	240		
QY	241	TGGAAGGATGCAAGTTATGAGTACCTCTGATGTGATGGTTCCTCCCAAGA	300		
Db	241	TGGAAGGATGCAAGTTATGAGTACCTCTGATGTGATGGTTCCTCCCAAGA	300		

241 TGGAAGGATGCAGGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
301 GATTTCAGAGCGAGACTTCAGGCAGACCTTCAGCGCTTTCCCTCATCGGATTCGCCAGCTA 360
301 GATTTCAGAGCGAGACTTCAGGCAGACCTTCAGCGCTTTCCCTCATCGGATTCGCCAGCTA 360
361 GCTAAATATGTTTCACAGCAAGGACTGAAGCTAGGGAATTTATGCAGATGTTGGAAATAAA 420
361 GCTAAATATGTTTCACAGCAAGGACTGAAGCTAGGGAATTTATGCAGATGTTGGAAATAAA 420
421 ACCTGCGCAGGCTTCCTCGGAGTTTGGATATCTACGACATGATGCCAGACCTTTGCT 480
421 ACCTGCGCAGGCTTCCTCGGAGTTTGGATATCTACGACATGATGCCAGACCTTTGCT 480
481 GACTGGGGAGTAGACTCTGCTAAATTTGATGGTTGTTACTGTGACAGCTTTGGAAATTTG 540
481 GACTGGGGAGTAGACTCTGCTAAATTTGATGGTTGTTACTGTGACAGCTTTGGAAATTTG 540
541 GCAGATGGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 600
541 GCAGATGGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 600
601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
661 CAGTACTGCMAATCAGCTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAGATATAAG 720
661 CAGTACTGCMAATCAGCTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAGATATAAG 720
721 AGTATCTTGAGCTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGACACAGG 780
721 AGTATCTTGAGCTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGACACAGG 780
781 GGTGGAAATGACCCAGATATGTTAGTGAATTTGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
781 GGTGGAAATGACCCAGATATGTTAGTGAATTTGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTTAATGACCTC 900
841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTTAATGACCTC 900
901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCATCAAT 960
901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCATCAAT 960
961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACTTTGAAGTGTG 1020
961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACTTTGAAGTGTG 1020
1021 GAACGACCTCTCAGGCTTAGCCTGGGCTGATGCTATGATTAACCGGACGAGATTTGTT 1080
1021 GAACGACCTCTCAGGCTTAGCCTGGGCTGATGCTATGATTAACCGGACGAGATTTGTT 1080
1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTCTAATCCT 1140
1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTCTAATCCT 1140
1141 GCCTGCTTCATCACAGCTCCTCCCTGTGAAAGGAGCTAGGGTTCTATGATGAGCT 1200
1141 GCCTGCTTCATCACAGCTCCTCCCTGTGAAAGGAGCTAGGGTTCTATGATGAGCT 1200
1201 TCAAGGTTAAGAAGTACATAAATCCACAGGACCTGTTTGTCTTCAAGCTATATA 1254
1201 TCAAGGTTAAGAAGTACATAAATCCACAGGACCTGTTTGTCTTCAAGCTATATA 1254

RESULT 2

US-10-103-327-15
; Sequence 15, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KIMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-10-103-327-15

Query Match 100.0%; Score 1254; DB 5; Length 1254;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCTTCGCTTCGCTTCCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCTTCGCTTCGCTTCCTGGCC 60
QY 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT 120
DB 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT 120
QY 121 ACCATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
DB 121 ACCATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 181 GATTCCTGCTCAGTGAGAGAGCTCTTCATGGAGATGGCAGAGCTCATGCTCTCAGAAGGC 240
DB 181 GATTCCTGCTCAGTGAGAGAGCTCTTCATGGAGATGGCAGAGCTCATGCTCTCAGAAGGC 240
QY 241 TGAAGGATGCAAGTTATGAGTACCTCTGCTGATGATGATGATGATGATGATGATGATGAT 300
DB 241 TGAAGGATGCAAGTTATGAGTACCTCTGCTGATGATGATGATGATGATGATGATGATGAT 300
QY 301 GATTCAGAGAGGAGAGCTTCAGGCAGACCTTCAGCGCTTTCCCTCATCGGATTCGCCAGCTA 360
DB 301 GATTCAGAGAGGAGAGCTTCAGGCAGACCTTCAGCGCTTTCCCTCATCGGATTCGCCAGCTA 360
QY 361 GCTAATTTATGTTTCACAGCAAGGACTGAAGCTAGGATTTATGCAATGTTGGAATAAA 420
DB 361 GCTAATTTATGTTTCACAGCAAGGACTGAAGCTAGGATTTATGCAATGTTGGAATAAA 420
QY 421 ACCTGGCGAGGCTTCCTCGGAGTTTGGATATCTACGACATGATGCCAGACCTTTGCT 480
DB 421 ACCTGGCGAGGCTTCCTCGGAGTTTGGATATCTACGACATGATGCCAGACCTTTGCT 480
QY 481 GACTGGGGAGTAGACTCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
DB 481 GACTGGGGAGTAGACTCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
QY 541 GCAGATGGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 600
DB 541 GCAGATGGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
QY 661 CAGTACTGCMAATCAGCTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAGATATAAG 720
DB 661 CAGTACTGCMAATCAGCTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAGATATAAG 720
QY 721 AGTATCTTGAGCTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGACACAGG 780

Db 721 AGTATCTTGACCTGAGCATCTTTTAACGAGAGAGAAATTTGATGTTGTCGACGAGG 780
Qy 781 GGTGGAATGACCCAGATATGTTAGTGATTTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAATGACCCAGATATGTTAGTGATTTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACCTCAGATGGCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTAATGACCTC 900
Qy 901 CGACATCAGCCTCAAGCCAAAGCTCTCTTCAGGTAAGGACGTAATGCGCAAT 960
Db 901 CGACATCAGCCTCAAGCCAAAGCTCTCTTCAGGTAAGGACGTAATGCGCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGATACCACTTTAGACAGGAGACAACTTTGAAGTGTG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGATACCACTTTAGACAGGAGACAACTTTGAAGTGTG 1020
Qy 1021 GAACGACCTCTCAGCCTTAGCCTGGCTGTAGCTATGATAAACCGGACGAGATTGGT 1080
Db 1021 GAACGACCTCTCAGCCTTAGCCTGGCTGTAGCTATGATAAACCGGACGAGATTGGT 1080
Qy 1081 GGACCTCGCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAATCCT 1140
Db 1081 GGACCTCGCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAATCCT 1140
Qy 1141 GCCTGCTTCATCACAGCTCCCTCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGAAT 1200
Db 1141 GCCTGCTTCATCACAGCTCCCTCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGAAT 1200
Qy 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254

RESULT 3

US-10-602-219-15
; Sequence 15, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grilli, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-219-15
Query Match 100.0%; Score 1254; DB 7; Length 1254;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCGCTTCCTCGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCGCTTCCTCGGCC 60
Qy 61 CTCGTTTCTCGGACATCTCCCTGGGCTAGAGCACTGGACAATGGATTTGGCAAGGACGCT 120
Db 61 CTCGTTTCTCGGACATCTCCCTGGGCTAGAGCACTGGACAATGGATTTGGCAAGGACGCT 120
Qy 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACTTGACTGCCAGGAGAGCCA 180
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACTTGACTGCCAGGAGAGCCA 180
Qy 181 GATTCTCGCATCAGTCAGGAAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAGGC 240
Db 181 GATTCTCGCATCAGTCAGGAAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAGGC 240
Qy 241 TGGAAGGATGCAAGTTATGAGTACTCTGCTCATTTGATGACTGTTGGATGGCTCCCCAAAGA 300
Db 241 TGGAAGGATGCAAGTTATGAGTACTCTGCTCATTTGATGACTGTTGGATGGCTCCCCAAAGA 300
Qy 301 GATTCAAGAGGACAGCTTCAGGACAGCCCTCAGGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTCAAGAGGACAGCTTCAGGACAGCCCTCAGGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATTTATGTTTACAGCAAGGAGCTGAAGCTAGGAGTTATGACAGATGTTGGAATATAA 420
Db 361 GCTAATTTATGTTTACAGCAAGGAGCTGAAGCTAGGAGTTATGACAGATGTTGGAATATAA 420
Qy 421 ACCTGCGAGGCTTCCCTGGGAGTTTGGATACTACGACATGATGATGCCAGAGCTTTGCT 480
Db 421 ACCTGCGAGGCTTCCCTGGGAGTTTGGATACTACGACATGATGATGCCAGAGCTTTGCT 480
Qy 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTACTGTGACAGATTGGAATTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTACTGTGACAGATTGGAATTTG 540
Qy 541 CGAGATGTTTATAAGCACATGCTTGGCCCTGTAATAGGACTGGCAGAGACTTGTGTAC 600
Db 541 CGAGATGTTTATAAGCACATGCTTGGCCCTGTAATAGGACTGGCAGAGACTTGTGTAC 600
Qy 601 TCCTGTGAGTGGCCTCTTTATATATGTCGCCCTTTCAAAGGCCCAATATATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCCTCTTTATATATGTCGCCCTTTCAAAGGCCCAATATATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCAGCTGGGCAAAATTTTGTGACATTTGATGATTCCTGGAAAAGTATAAAG 720
Db 661 CAGTACTGCAATCAGCTGGGCAAAATTTTGTGACATTTGATGATTCCTGGAAAAGTATAAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGACAGAGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGACAGAGG 780
Qy 781 GGTGGAATGACCCAGATATGTTAGTATTTGGCAACTTTGGGCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAATGACCCAGATATGTTAGTATTTGGCAACTTTGGGCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACCTCAGATGGCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTAATGACCTC 900
Qy 901 CGACATCAGCCTCAAGCCAAAGCTCTCTTCAGGTAAGGACGTAATGCGCAAT 960
Db 901 CGACATCAGCCTCAAGCCAAAGCTCTCTTCAGGTAAGGACGTAATGCGCAAT 960

QY 961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTAGACAGGAGACAACTTTGAAGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTAGACAGGAGACAACTTTGAAGTGG 1020
QY 1021 GAACACCTCTCAGGCTTACGCTGGGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080
DB 1021 GAACACCTCTCAGGCTTACGCTGGGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080
QY 1081 GGACCTCGCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAACTCT 1140
DB 1081 GGACCTCGCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAACTCT 1140
QY 1141 GCCTGCTTCATCACAGCTCCTCCTGTGTAAGAGAACTAGGCTTCTATGAATGGACT 1200
DB 1141 GCCTGCTTCATCACAGCTCCTCCTGTGTAAGAGAACTAGGCTTCTATGAATGGACT 1200
QY 1201 TCAGGTTAAGAGTACATAAATCCACAGGCACTGTTTTCCTCAGCTATAA 1254
DB 1201 TCAGGTTAAGAGTACATAAATCCACAGGCACTGTTTTCCTCAGCTATAA 1254

RESULT 4

US-10-602-220-15
; Sequence 15, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP07B
; CURRENT APPLICATION NUMBER: US/10/602,220
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-220-15

Query Match 100.0%; Score 1254; DB 7; Length 1254;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATCGAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCCTGGCC 60
DB 1 ATCGAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCCTGGCC 60

QY 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTTGCAAGGAGCGCT 120
DB 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTTGCAAGGAGCGCT 120
QY 121 ACCATGGGCTGCTGCATCTGGGAGCGCTTCATGTGCAACCTTTGACTGCGCAGGAAGGCCA 180
DB 121 ACCATGGGCTGCTGCATCTGGGAGCGCTTCATGTGCAACCTTTGACTGCGCAGGAAGGCCA 180
QY 181 GATTCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCTCAGAGGC 240
DB 181 GATTCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCTCAGAGGC 240
QY 241 TCGAAGGATGCAAGTTATGAGTACCTCTGCATTTGATGACTGTTGAGATGGCTCCCCAAGA 300
DB 241 TCGAAGGATGCAAGTTATGAGTACCTCTGCATTTGATGACTGTTGAGATGGCTCCCCAAGA 300
QY 301 GATTCTGAGAGGAGAGCTTCAGGCAGACCTTCAGCGCTTTCTCATGGGATTCGCCAGCTA 360
DB 301 GATTCTGAGAGGAGAGCTTCAGGCAGACCTTCAGCGCTTTCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATTATTTACAGCAAGGAGCTGAAGCTAGGATTTATGACAGATGTTGGAATAAA 420
DB 361 GCTAATTATTTACAGCAAGGAGCTGAAGCTAGGATTTATGACAGATGTTGGAATAAA 420
QY 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTCGATGCCAGACCTTTTGTCT 480
DB 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTCGATGCCAGACCTTTTGTCT 480
QY 481 GACTGGGAGTAGATCTGCTAAAAATTTGATGTTGTTACTGTGACAGATTTGGAAAAATTTG 540
DB 481 GACTGGGAGTAGATCTGCTAAAAATTTGATGTTGTTACTGTGACAGATTTGGAAAAATTTG 540
QY 541 GCAGATGTTATAAGCAGATGCTTGGCCCTGAATAGACTGGCAGAGACATTTGTGTAC 600
DB 541 GCAGATGTTATAAGCAGATGCTTGGCCCTGAATAGACTGGCAGAGACATTTGTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCAGCTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAAATTAAG 720
DB 661 CAGTACTGCAATCAGCTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAAATTAAG 720
QY 721 AGTATCTTGCATGAGCACTTTTAAACAGGAGAGATTTGTCATGCTGTCGACAGG 780
DB 721 AGTATCTTGCATGAGCACTTTTAAACAGGAGAGATTTGTCATGCTGTCGACAGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTATGTCGAACTTTGGCCCTCAGCTGGAATCAGCAA 840
DB 781 GGTGGAATGACCCAGATATGTTAGTATGTCGAACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
DB 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
QY 901 CGACATCATGCCCCTCAAGCCAAAGCTCTCTTTCAGATAAGCAGCGTAATTTGCCATCAAT 960
DB 901 CGACATCATGCCCCTCAAGCCAAAGCTCTCTTTCAGATAAGCAGCGTAATTTGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACCGCTTTAGCAGGAGAGCACTTTGAAGTGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGTTACCGCTTTAGCAGGAGAGCACTTTGAAGTGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTGGGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTGGGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080
QY 1081 GGAACCTGCTCTTATACCAATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAACTCT 1140
DB 1081 GGAACCTGCTCTTATACCAATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAACTCT 1140

QY 841 GTAACATCAGATGCCCTCTGGGCTATCATGGCTGCTCCCTTTATTCATGCTTAATGACCTC 900
DB |||||||GTAACATCAGATGCCCTCTGGGCTATCATGGCTGCTCCCTTTATTCATGCTTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTGAGGATAGGACGTAATTTGCCATCAAT 960
DB |||||||CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTGAGGATAGGACGTAATTTGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGGTACCGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
DB |||||||CAGGACCCCTTGGGCAAGCAAGGGTACCGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGGCTGAGCTGAGCTATGATAAACCAGGAGAGATTGGT 1080
DB |||||||GAACGACCTCTCTCAGGCTTAGGCTGAGCTGAGCTATGATAAACCAGGAGAGATTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAATCCT 1140
DB |||||||GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAATCCT 1140
QY 1141 GCTGCTTCATACACAGCTCCCTCTGTGAAAGAGAGCTAGGGTTCTATGATGACT 1200
DB |||||||GCTGCTTCATACACAGCTCCCTCTGTGAAAGAGAGCTAGGGTTCTATGATGACT 1200
QY 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTTCAGCTATATA 1254
DB |||||||TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTTCAGCTATATA 1254

RESULT 9
US-10-602-219-11
; Sequence 11, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent In version 3.2
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-219-11

Query Match 99.9%; Score 1252.4; DB 7; Length 1266;

Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGNAACCCAGAACTACATCTGGGCTGCGCGCTTGCCTTCCTTCCTGGCC 60
DB 1 ATGCAGCTGAGNAACCCAGAACTACATCTGGGCTGCGCGCTTGCCTTCCTTCCTGGCC 60
QY 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTTGGCAAGAGCGCT 120
DB 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTTGGCAAGAGCGCT 120
QY 121 ACCATGGCTGGCTGCACTGGGAGCGCTTTCATGTGCAACCTTGCACGCGGAGAGGCCA 180
DB 121 ACCATGGCTGGCTGCACTGGGAGCGCTTTCATGTGCAACCTTGCACGCGGAGAGGCCA 180
QY 181 GATTCTGTCATGATGAGAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAGGC 240
DB 181 GATTCTGTCATGATGAGAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAGGC 240
QY 241 TGGAGGATGTCAGGTTTATGAGTACCTCTGATTGATGACTGTTGGATGGCTCCCAAAGA 300
DB 241 TGGAGGATGTCAGGTTTATGAGTACCTCTGATTGATGACTGTTGGATGGCTCCCAAAGA 300
QY 301 GATTCAAGAGGCAAGCTTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
DB 301 GATTCAAGAGGCAAGCTTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATTTGTTACAGCAAGAGCTGAAAGCTAGGATTTATGCAGATGTTGGAATATAA 420
DB 361 GCTAATTTGTTACAGCAAGAGCTGAAAGCTAGGATTTATGCAGATGTTGGAATATAA 420
QY 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATGATGATGCCAGACCTTTGCT 480
DB 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATGATGATGCCAGACCTTTGCT 480
QY 481 GACTGGGGAGTAGATCTGCTTAAATTTGATGGTTGATCTGTCACAGTTTGGAAAAATTG 540
DB 481 GACTGGGGAGTAGATCTGCTTAAATTTGATGGTTGATCTGTCACAGTTTGGAAAAATTG 540
QY 541 GCAGATGTTTAAAGCAGATGCTTGGCCCTCAATAGGACTGCGAGAGCATTTGTGTAC 600
DB 541 GCAGATGTTTAAAGCAGATGCTTGGCCCTCAATAGGACTGCGAGAGCATTTGTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAAATTTTGTGCACTTGTGATGATTCCTGGAAAAAGTATAAG 720
DB 661 CAGTACTGCAATCACTGGCGAAAATTTTGTGCACTTGTGATGATTCCTGGAAAAAGTATAAG 720
QY 721 AGTATCTTGGACTGGACATCTTTTAAACCAGGAGAGAAATTTGATGTTGCTGGACCAAGG 780
DB 721 AGTATCTTGGACTGGACATCTTTTAAACCAGGAGAGAAATTTGATGTTGCTGGACCAAGG 780
QY 781 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCTCAGCTGGAATCAGCAA 840
DB 781 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCTCAGCTGGAATCAGCAA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTGAGGATAGGACGTAATTTGCCATCAAT 960
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTGAGGATAGGACGTAATTTGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGGTACCGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGGTACCGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGGCTGAGCTGAGCTATGATAAACCAGGAGAGATTGGT 1080
DB |||||||GAACGACCTCTCTCAGGCTTAGGCTGAGCTGAGCTATGATAAACCAGGAGAGATTGGT 1080

Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCAGGAGGAGTTGGT 1080
Qy 1081 GGACCTCGCTCTTATACATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Db 1081 GGACCTCGCTCTTATACATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Qy 1141 GCCTGCTTCATCACAGAGCTCCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGATGAGCT 1200
Db 1141 GCCTGCTTCATCACAGAGCTCCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGATGAGCT 1200
Qy 1201 TCAAGGTTAAGAAAGTACATAAATCCACAGGCACTGTTTTGCTTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAAAGTACATAAATCCACAGGCACTGTTTTGCTTCAGCTAGAA 1254

RESULT 10

US-10-602-220-11
; Sequence 11, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION

; FILE REFERENCE: LSBC-0087-CP07B
; CURRENT APPLICATION NUMBER: US/10/602,220
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-220-11

Query Match 99.9%; Score 1252.4; DB 7; Length 1266;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTGGCTTCCCTGGCC 60
Db 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTGGCTTCCCTGGCC 60
Qy 61 CTCGTTTCTGGACATCCCTGGGCTTAGAGCACTGACAAATGGAATGGCAAGGACGCCT 120
Db 61 CTCGTTTCTGGACATCCCTGGGCTTAGAGCACTGACAAATGGAATGGCAAGGACGCCT 120
Qy 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCACCTTGACCTGCCAGGAGGCCA 180

Db 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCACCTTGACCTGCCAGGAGGCCA 180
Qy 181 GATTCTTCATAGTGAAGAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
Db 181 GATTCTTCATAGTGAAGAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
Qy 241 TGGAAAGATGACAGTTATGAGTACCTCTGCATTTGATGACTGTTGGATGGCTCCCAAGA 300
Db 241 TGGAAAGATGACAGTTATGAGTACCTCTGCATTTGATGACTGTTGGATGGCTCCCAAGA 300
Qy 301 GATTTCAGAAAGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTTCAGAAAGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATTTATGTTTCAAGCAAGGACTGAAAGCTAGGAGTTTATCAGAGATGTTGAAATATAA 420
Db 361 GCTAATTTATGTTTCAAGCAAGGACTGAAAGCTAGGAGTTTATCAGAGATGTTGAAATATAA 420
Qy 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACATCAGACATTCATGCCAGACCTTTGCT 480
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACATCAGACATTCATGCCAGACCTTTGCT 480
Qy 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTTACTGTGACAGTTTGGAAATTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTTACTGTGACAGTTTGGAAATTTG 540
Qy 541 GCAGATGGTTTATAAGCACATGCTCTTGGCCCTGAAATAGGACTGGCAGAGCATTTGTATC 600
Db 541 GCAGATGGTTTATAAGCACATGCTCTTGGCCCTGAAATAGGACTGGCAGAGCATTTGTATC 600
Qy 601 TCCTGTGAGTGGCTCTTTATATGTCGGCCCTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATGTCGGCCCTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCAGTGGCGAAATTTTGTGCAATTTGATGATGATTCCTGGAAAGTATAAAG 720
Db 661 CAGTACTGCAATCAGTGGCGAAATTTTGTGCAATTTGATGATGATTCCTGGAAAGTATAAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTTGTGATGTTGTCGACCAAGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTTGTGATGTTGTCGACCAAGG 780
Qy 781 GGTGGAATGACCCAGATATGTTAGTGTGGCACTTTGGCCCTCAGCTGGGAATCAGCAA 840
Db 781 GGTGGAATGACCCAGATATGTTAGTGTGGCACTTTGGCCCTCAGCTGGGAATCAGCAA 840
Qy 841 GTAACTCAGATGSCCTCTGGGCTATCATGGCTGCTCTTTTATTTCATGCTTAATGACCTC 900
Db 841 GTAACTCAGATGSCCTCTGGGCTATCATGGCTGCTCTCTTTTATTTCATGCTTAATGACCTC 900
Qy 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCGCAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCGCAGGAGATTGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCCTGTAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCCTGTAATCCT 1140
Qy 1141 GCCTGCTTCATCACAGCTCCTCCTGTGTAAGGAAGTAGGGTTCTATGATGAGACT 1200
Db 1141 GCCTGCTTCATCACAGCTCCTCCTGTGTAAGGAAGTAGGGTTCTATGATGAGACT 1200
Qy 1201 TCAAGGTTAAGAAAGTACATAAATCCACAGGCACTGTTTTGCTTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAAAGTACATAAATCCACAGGCACTGTTTTGCTTCAGCTAGAA 1254

```
RESULT 11
US-10-851-388-11
; Sequence 11, Application US/10851388
; Publication No. US20040234516A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851,388
; PRIOR FILING DATE: 2004-05-21
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-10-851-388-11
Query Match 99.9%; Score 1252.4; DB 8; Length 1266;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCCAGCTGAGAACCCAGAACTACATCTGGGCTGCGCGCTTCCGCTTCCGCTTCCGCGC 60
Db 1 ATCCAGCTGAGAACCCAGAACTACATCTGGGCTGCGCGCTTCCGCTTCCGCTTCCGCGC 60
QY 61 CTGGTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGGATGGCAAGCGCCT 120
Db 61 CTGGTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGGATGGCAAGCGCCT 120
QY 121 ACCATGGGCTGGCTGCATCGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGGCCA 180
Db 121 ACCATGGGCTGGCTGCATCGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGGCCA 180
QY 181 GATTCTGTCATAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Db 181 GATTCTGTCATAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
QY 241 TGGAGGATGCAGGTTATGATGACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Db 241 TGGAGGATGCAGGTTATGATGACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
QY 301 GATTCAAGGCGAGACTTCAGGCGAGACCTCCTCAGCGCTTTCCTCATGGATTGCCAGCTA 360
Db 301 GATTCAAGGCGAGACTTCAGGCGAGACCTCCTCAGCGCTTTCCTCATGGATTGCCAGCTA 360
QY 361 GCTAAATATGTTACAGCAAGGACTGAAGTGGATTTATGAGATTTGGAATAAA 420
Db 361 GCTAAATATGTTACAGCAAGGACTGAAGTGGATTTATGAGATTTGGAATAAA 420
QY 421 ACCTGGCGGAGCTTCCCTGGGAGTTTGGATACCTACGACATGATGCCAGAGCTTTGCT 480
Db 421 ACCTGGCGGAGCTTCCCTGGGAGTTTGGATACCTACGACATGATGCCAGAGCTTTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTTGTACTGTGACAGTTTGGAAAATTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTTGTACTGTGACAGTTTGGAAAATTTG 540
QY 541 GCAGATGGTTATAGCAATGCTTCCCTGGCCCTGAAATAGGACTGGCAGAGCATTTGTGTAC 600
Db 541 GCAGATGGTTATAGCAATGCTTCCCTGGCCCTGAAATAGGACTGGCAGAGCATTTGTGTAC 600
QY 601 TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATGATGATTCCTGGGAAAGATATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATGATGATTCCTGGGAAAGATATAAG 720
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTGATGCTGCTGACAGGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTGATGCTGCTGACAGGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTATGGCACTTTGGGCTCAGCTGGAAATCAGCAA 840
Db 781 GGTGGAATGACCCAGATATGTTAGTATGGCACTTTGGGCTCAGCTGGAAATCAGCAA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGCTAATTCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGCTAATTCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGAGCTATCATATAACCGGAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGAGCTATCATATAACCGGAGGAGATTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTCTAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTCTAATCCT 1140
QY 1141 GCCTGCTTCATCACAGCTCCTCCTGTGAAAGAAAGTAGGGTTCTATGATGAGCT 1200
Db 1141 GCCTGCTTCATCACAGCTCCTCCTGTGAAAGAAAGTAGGGTTCTATGATGAGCT 1200
QY 1201 TCAAGTTAAGAAGTCAATTAATCCACAGGACCTGTTTGTCTTCAGCTATAA 1254
Db 1201 TCAAGTTAAGAAGTCAATTAATCCACAGGACCTGTTTGTCTTCAGCTATAA 1254
;
; RESULT 12
; US-10-984-389-11
; Sequence 11, Application US/10984389
; Publication No. US20050125859A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/984,389
; CURRENT FILING DATE: 2004-11-08
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-10-984-389-11
Query Match 99.9%; Score 1252.4; DB 9; Length 1266;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCCTGCGCC 60
Db 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCCTGCGCC 60
Qy 61 CTGGTTTCTGGGACATCCCTGGGCTAGAGCACTGAGCAATGGATGGCAAGAGCGCT 120
Db 61 CTGGTTTCTGGGACATCCCTGGGCTAGAGCACTGAGCAATGGATGGCAAGAGCGCT 120
Qy 121 ACCATGGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
Db 121 ACCATGGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
Qy 181 GATTCTGCACTAGTGAGAGAGCTTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAAGC 240
Db 181 GATTCTGCACTAGTGAGAGAGCTTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAAGC 240
Qy 241 TGAAGAGATCAGGTATAGTACTCTCTGATGATGATGATGATGATGATGATGATGATGAT 300
Db 241 TGAAGAGATCAGGTATAGTACTCTCTGATGATGATGATGATGATGATGATGATGATGAT 300
Qy 301 GATTCAAGGAGAGCTTCAAGGAGAGCTTCAAGGAGAGCTTCAAGGAGAGCTTCAAGGAG 360
Db 301 GATTCAAGGAGAGCTTCAAGGAGAGCTTCAAGGAGAGCTTCAAGGAGAGCTTCAAGGAG 360
Qy 361 GCTAAATATGTTTCAAGCAAGGAGCTGAAGCTAGGAGATTTATGAGATGTTGGAATAAAA 420
Db 361 GCTAAATATGTTTCAAGCAAGGAGCTGAAGCTAGGAGATTTATGAGATGTTGGAATAAAA 420
Qy 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATACCTAGCAATGATGATGATGATGATGATGAT 480
Db 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATACCTAGCAATGATGATGATGATGATGATGAT 480
Qy 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTTGTACTGTGACAGTTTGGAAATTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTTGTACTGTGACAGTTTGGAAATTTG 540
Qy 541 GCAGATGTTTAAAGCAATGTCCTGGGCTGAAATAGGAGCTGGCAAGCAATTTGTATC 600
Db 541 GCAGATGTTTAAAGCAATGTCCTGGGCTGAAATAGGAGCTGGCAAGCAATTTGTATC 600
Qy 601 TCTGTGAGTGCTCTTTTATATGTCGCTTTCCTGAAAGCCCAATTTATAGAAATCCGA 660
Db 601 TCTGTGAGTGCTCTTTTATATGTCGCTTTCCTGAAAGCCCAATTTATAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCAATGATGATGATGATGATGATGATGATGAT 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCAATGATGATGATGATGATGATGATGATGAT 720
Qy 721 AGTATCTTGAATGAGCAATCTTTTAAACAGGAGAGATTTGATGTTGCTGACCAAGG 780
Db 721 AGTATCTTGAATGAGCAATCTTTTAAACAGGAGAGATTTGATGTTGCTGACCAAGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGCAATCTTGGCTCAGCTGGATCAGCAA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGCAATCTTGGCTCAGCTGGATCAGCAA 840
Qy 841 GTAACTCAGATGCTCTGGCTTATCATGCTGCTCTTTTATTTATGTTCTAATGACCTC 900
Db 841 GTAACTCAGATGCTCTGGCTTATCATGCTGCTCTTTTATTTATGTTCTAATGACCTC 900
Qy 901 CGACACATCAGCCCTCAAGCAAGCTCTCTTTTCAAGGATAGGAGCGTAATTTGCCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCAAGCTCTCTTTTCAAGGATAGGAGCGTAATTTGCCATCAAT 960
Qy 961 CAGGACCCCTTGGCAAGAGGATACCACTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGCAAGAGGATACCACTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGTATGATTAACCGGAGGAGATTTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGTATGATTAACCGGAGGAGATTTGGT 1080
```

```
Qy 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Qy 1141 GCCTGCTTTCATCACAGAGCTCCCTGTCGAAAGGAAGCTAGGGTTCTATGATGAGCT 1200
Db 1141 GCCTGCTTTCATCACAGAGCTCCCTGTCGAAAGGAAGCTAGGGTTCTATGATGAGCT 1200
Qy 1201 TCAAGGTTAAGAAAGTCAATAAAATCCACAGGCACTGTTTTGCTTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAAAGTCAATAAAATCCACAGGCACTGTTTTGCTTCAGCTAGAA 1254
```

```
RESULT 13
US-09-993-059-7
; Sequence 7, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-09-993-059-7
```

```
Query Match 99.9%; Score 1252.4; DB 3; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCCTGCGCC 60
Db 1 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCCTGCGCC 60
Qy 61 CTGGTTTCTGGGACATCCCTGGGCTAGAGCACTGAGCAATGGATGGCAAGAGAGCGCT 120
Db 61 CTGGTTTCTGGGACATCCCTGGGCTAGAGCACTGAGCAATGGATGGCAAGAGAGCGCT 120
Qy 121 ACCATGGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
Db 121 ACCATGGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
Qy 181 GATTCTGCACTAGTGAGAGAGCTTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAAGC 240
Db 181 GATTCTGCACTAGTGAGAGAGCTTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAAGC 240
Qy 241 TGAAGAGATCAGGTATAGTACTCTCTGATGATGATGATGATGATGATGATGATGATGAT 300
Db 241 TGAAGAGATCAGGTATAGTACTCTCTGATGATGATGATGATGATGATGATGATGATGAT 300
Qy 301 GATTCAAGGAGAGCTTCAAGGAGAGCTTCAAGGAGAGCTTCAAGGAGAGCTTCAAGGAG 360
Db 301 GATTCAAGGAGAGCTTCAAGGAGAGCTTCAAGGAGAGCTTCAAGGAGAGCTTCAAGGAG 360
Qy 361 GCTAAATATGTTTCAAGCAAGGAGCTGAAGCTAGGAGATTTATGAGATGTTGGAATAAAA 420
Db 361 GCTAAATATGTTTCAAGCAAGGAGCTGAAGCTAGGAGATTTATGAGATGTTGGAATAAAA 420
Qy 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATACCTAGCAATGATGATGATGATGATGATGAT 480
Db 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATACCTAGCAATGATGATGATGATGATGATGAT 480
```


Qy 961 CAGGACCCCTTGGGCAAGGCTACAGCTTATAGACAGGAGACAACTTTGAAGTGG 1020
Db CAGGACCCCTTGGGCAAGGCTACAGCTTATAGACAGGAGACAACTTTGAAGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAACCGCAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAACCGCAGGAGATTGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCTGTAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCTGTAATCCT 1140
Qy 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGTAAGAGAGTGGCTGTAATGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGTAAGAGAGTGGCTGTAATGACT 1200
Qy 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGTCTTACAGCTATA 1254
Db 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGTCTTACAGCTATA 1254

RESULT 15

US-10-602-219-7
; Sequence 7, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: US/10/602,219
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-219-7

Query Match 99.98; Score 1252.4; DB 7; Length 1278;
Best Local Similarity 99.98; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTGGCTTCTTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTGGCTTCTTGGCC 60

Qy 61 CTCGTTTCTGGGACATCCCTGGGGCTAGAGCAATGGAATGGATGGCAAGGACGCT 120
Db CTCGTTTCTGGGACATCCCTGGGGCTAGAGCAATGGAATGGATGGCAAGGACGCT 120
Qy 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Qy 181 GATTTCCTGCATCAGTGAAGAGCTCTTCATGAGAGTGGCAGAGCTCATGGTCTTCAGAAAGC 240
Db 181 GATTTCCTGCATCAGTGAAGAGCTCTTCATGAGAGTGGCAGAGCTCATGGTCTTCAGAAAGC 240
Qy 241 TGAAGAGTGCAGGTTATGAGTACCTCTGCAATGATGACTTGGTGGATGGCTCCCCAAGA 300
Db 241 TGAAGAGTGCAGGTTATGAGTACCTCTGCAATGATGACTTGGTGGATGGCTCCCCAAGA 300
Qy 301 GATTTCAGAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCGAGCTA 360
Db 301 GATTTCAGAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCGAGCTA 360
Qy 361 GCTAATTTATGTTTCAAGCAAGGAGCTGAAGCTAGGGATTTATGCAGATTTGGAAATATA 420
Db 361 GCTAATTTATGTTTCAAGCAAGGAGCTGAAGCTAGGGATTTATGCAGATTTGGAAATATA 420
Qy 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATGATGATGCCAGACCTTTGCT 480
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATGATGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGGAGTAGACTCTGCTTAAATTTGATGGTTTGTACTGTGCACAGTTGGGAAATTTG 540
Db 481 GACTGGGGAGTAGACTCTGCTTAAATTTGATGGTTTGTACTGTGCACAGTTGGGAAATTTG 540
Qy 541 GCAGATGGTTTAAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAAAGCTTTGTGCTAC 600
Db 541 GCAGATGGTTTAAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAAAGCTTTGTGCTAC 600
Qy 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTCAAAGGCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTCAAAGGCCAATTTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCAGCTGGGAAATTTTGTGACATGATGATTCCTGGAAAGTATATAAG 720
Db 661 CAGTACTGCAATCAGCTGGGAAATTTTGTGACATGATGATTCCTGGAAAGTATATAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTTGATGTTGCTGGACGAGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTTGATGTTGCTGGACGAGG 780
Qy 781 GGTGGAATGACCCAGATATGTTAGTATGTCGCACTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAATGACCCAGATATGTTAGTATGTCGCACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAACCGCAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAACCGCAGGAGATTGGT 1080
Qy 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCTTGAATCCT 1140
Db 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCTTGAATCCT 1140
Qy 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTGTAAGAGAGTGGCTTCTATGATGAGTACT 1200

Db 1141 GCTGCTTCATCACAGCTCCTCCCTGTAAGGAGGAGCTAGGGTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTGCTTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTGCTTCAGCTAGAA 1254

Search completed: December 31, 2005, 23:36:16
Job time : 1165 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 31, 2005, 21:00:36 ; Search time 247 Seconds
(without alignments)
2661.468 Million cell updates/sec

Title: US-10-602-220-15
Perfect score: 1254
Sequence: 1 atgcagctgagggaaccaga.....ctgttttgcagctataa 1254

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4172979 seqs, 262114271 residues

Total number of hits satisfying chosen parameters: 8345958

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA New.*
1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2.*
9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3.*
10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1252.4	99.9	1355	US-10-981-267-26	Sequence 26, Appl
2	1252.4	99.9	1547	US-10-981-267-25	Sequence 25, Appl
3	37	3.0	2779	US-10-750-185-40681	Sequence 40681, A
4	36.6	2.9	1306	US-10-750-185-42255	Sequence 42255, A
5	35	2.8	321019	US-10-995-561-13204	Sequence 13204, A
6	34	2.7	1932	US-10-750-185-55363	Sequence 55363, A
7	33	2.6	1437	US-10-750-185-43877	Sequence 43877, A
8	32.8	2.6	1666	US-10-821-234-149	Sequence 149, App
9	32.8	2.6	146658	US-11-121-086-68	Sequence 68, Appl
10	32.8	2.6	188682	US-11-112-908-23	Sequence 23, Appl
11	32.8	2.6	190882	US-11-121-086-69	Sequence 69, Appl
12	32.4	2.6	1685	US-10-750-185-53897	Sequence 53897, A
13	32.4	2.6	86081	US-10-995-561-13246	Sequence 13246, A
14	32	2.6	1254	US-10-750-185-29504	Sequence 29504, A
15	31.8	2.5	201	US-10-995-561-29100	Sequence 29100, A
16	31.8	2.5	4847	US-10-750-185-60280	Sequence 60280, A
17	31.6	2.5	2787	US-10-841-129-3	Sequence 3, Appl
18	31.6	2.5	1125000	US-10-995-561-13286	Sequence 13286, A
19	31.4	2.5	119160	US-11-121-086-12	Sequence 12, Appl
20	31.4	2.5	1161874	US-11-121-086-75	Sequence 75, Appl
21	31.2	2.5	1522	US-10-750-185-37196	Sequence 37196, A
22	31	2.5	854	US-10-750-185-28311	Sequence 28311, A
23	31	2.5	1084	US-10-750-185-61883	Sequence 61883, A

Sequence 37071, A
Sequence 3, Appl
Sequence 64685, A
Sequence 121, App
Sequence 357, App
Sequence 54302, A
Sequence 13493, A
Sequence 42925, A
Sequence 66, Appl
Sequence 60, Appl
Sequence 39400, A
Sequence 1, Appl
Sequence 13227, A
Sequence 43163, A
Sequence 28504, A
Sequence 39271, A
Sequence 35062, A
Sequence 63028, A
Sequence 4, Appl
Sequence 5, Appl
Sequence 2, Appl
Sequence 6, Appl

US-10-750-185-37071
US-11-121-086-3
US-10-750-185-64685
US-10-131-826A-121
US-10-821-234-357
US-10-750-185-54302
US-10-995-561-13493
US-10-750-185-42925
US-11-121-086-66
US-11-121-086-60
US-10-750-185-39400
US-11-121-086-1
US-10-995-561-13227
US-10-750-185-43163
US-10-750-185-28504
US-10-750-185-39271
US-10-750-185-35062
US-10-750-185-63028
US-10-509-921-4
US-10-509-921-5
US-10-509-921-2
US-10-509-921-6

ALIGNMENTS

RESULT 1

US-10-981-267-26
; Sequence 26, Application US/10981267
; Publication No. US20050281805A1
; GENERAL INFORMATION:
; APPLICANT: LeBowitz, Jonathan H
; APPLICANT: Beverley, Stephen
; APPLICANT: Sly, William S.
; TITLE OF INVENTION: TARGETED THERAPEUTIC PROTEINS
; FILE REFERENCE: SYM-009CP
; CURRENT APPLICATION NUMBER: US/10/981,267
; CURRENT FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/384,452
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: US 60/516,990
; PRIOR FILING DATE: 2003-11-03
; PRIOR APPLICATION NUMBER: PCT/US03/17211
; PRIOR FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: US 10/272,531
; PRIOR FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US 60/445,734
; PRIOR FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: US 60/386,019
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: US 60/408,816
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 26
; LENGTH: 1355
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-981-267-26

Query Match 99.9%; Score 1252.4; DB 6; Length 1355;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAGACCAGACTACATCTGGGCTGGCGCTTCGCTTCCTTCCTGGCC 60
Db 48 ATGCAGCTGAGGAGACCAGACTACATCTGGGCTGGCGCTTCGCTTCCTTCCTGGCC 107
QY 61 CTCGTTTCTGGGACATCCTGGGGCTAGAGCACTGGCAATGGATTGGCAAGACGCCT 120
Db 108 CTCGTTTCTGGGACATCCTGGGGCTAGAGCACTGGCAATGGATTGGCAAGACGCCT 167

121 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
168 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 227
181 GATTCTGCATCAGTGAAGACTCTTCATGGAGATGGCAGAGCTCATGFTCTCAGAAGGC 240
228 GATTCTGCATCAGTGAAGACTCTTCATGGAGATGGCAGAGCTCATGFTCTCAGAAGGC 287
241 TGAAGGATGCAGGTTATGAGTACCTCTGATGATGACTGTTGGATGGCTCCCAAGA 300
288 TGAAGGATGCAGGTTATGAGTACCTCTGATGATGACTGTTGGATGGCTCCCAAGA 347
301 GATTCAAGAGCAGACTTCAGGCAGACCTTCAGCGCTTCTCATGGATTCGCCAGCTA 360
348 GATTCAAGAGCAGACTTCAGGCAGACCTTCAGCGCTTCTCATGGATTCGCCAGCTA 407
361 GCTAATTATGTTACAGCAAGAGCTGAAGCTAGGATTTATCAGATGTTGGAATAAA 420
408 GCTAATTATGTTACAGCAAGAGCTGAAGCTAGGATTTATCAGATGTTGGAATAAA 467
421 ACCTGGCAGGCTTCCTGGAGTTTGGATCTACGACATTCATGCCAGACCTTTGCT 480
468 ACCTGGCAGGCTTCCTGGAGTTTGGATCTACGACATTCATGCCAGACCTTTGCT 527
481 GACTGGGAGCTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTG 540
528 GACTGGGAGTATCTGCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTG 587
541 GCAGATGGTTATAGCACATGCTCTGGCCCTGAATAGGACTGGCAGAGCATTTGTATC 600
588 GCAGATGGTTATAGCACATGCTCTGGCCCTGAATAGGACTGGCAGAGCATTTGTATC 647
601 TCCTGTAGTGGCTCTTTATATATGTCCTTCAAAGCCCAATTATACAGAAATCCGA 660
648 TCCTGTAGTGGCTCTTTATATATGTCCTTCAAAGCCCAATTATACAGAAATCCGA 707
661 CAGTACTGCAATCACTGGCGAAATTTGCTGACATGATGATCTCTGGAAAGTATAAAG 720
708 CAGTACTGCAATCACTGGCGAAATTTGCTGACATGATGATCTCTGGAAAGTATAAAG 767
721 AGTATCTTGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGTCGACAGGG 780
768 AGTATCTTGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGTCGACAGGG 827
781 GGTGGAAATGACCCAGATATGTTAGTATGTTGGCACTTTGGCCCTCAGCTGGAATCAGCA 840
828 GGTGGAAATGACCCAGATATGTTAGTATGTTGGCACTTTGGCCCTCAGCTGGAATCAGCA 887
841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTATGTTCTAATGACCTC 900
888 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTATGTTCTAATGACCTC 947
901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAAGGACGTAATTCATCAAT 960
948 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAAGGACGTAATTCATCAAT 1007
961 CAGAACCCCTTGGCAAGAGGGTACAGCTTAGACAGGAGACAACTTTGAAGTGG 1020
1008 CAGAACCCCTTGGCAAGAGGGTACAGCTTAGACAGGAGACAACTTTGAAGTGG 1067
1021 GAACGACCTCTCAGGCTTAGCTGGCTGTAGCTATGATTAACCGGAGGAGATGGT 1080
1068 GAACGACCTCTCAGGCTTAGCTGGCTGTAGCTATGATTAACCGGAGGAGATGGT 1127
1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
1128 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1187
1141 GCCTGCTTATCAGACGCTCTCCCTGTGAAGAGAGCTAGGGTCTATGATGAGCT 1200
1188 GCCTGCTTATCAGACGCTCTCCCTGTGAAGAGAGCTAGGGTCTATGATGAGCT 1247
1201 TCAAGGTTAAGAGTCAATAAATCCCAAGGCACTGTTTGTCTTCAGCTATAA 1254

Db 1248 TCAAGGTTAAGAGTCAATAAATCCCAAGGCACTGTTTGTCTTCAGCTAGAA 1301
RESULT 2
US-10-981-267-25
; Sequence 25, Application US/10981267
; Publication No. US20050281805A1
; GENERAL INFORMATION:
; APPLICANT: Lebowitz, Jonathan H
; APPLICANT: Beverley, Stephen
; APPLICANT: Sly, William S.
; TITLE OF INVENTION: TARGETED THERAPEUTIC PROTEINS
; FILE REFERENCE: SYM-009CP
; CURRENT APPLICATION NUMBER: US/10/981,267
; CURRENT FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/384,452
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: US 60/516,990
; PRIOR FILING DATE: 2003-11-03
; PRIOR APPLICATION NUMBER: PCT/US03/17211
; PRIOR FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: US 10/272,531
; PRIOR FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US 60/445,734
; PRIOR FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: US 60/386,019
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: US 60/408,816
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 25
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: An exemplary GILT-tagged alpha-GAL A cassette sequence
US-10-981-267-25
Query Match 99.9%; Score 1252.4; DB 6; Length 1547;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTTCCTGCGCC 60
Db 51 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTTCCTGCGCC 110
Qy 61 CTCGTTTCTGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGACGCT 120
Db 111 CTCGTTTCTGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGACGCT 170
Qy 121 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Db 171 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 230
Qy 181 GATTCTTCATCAGTGAAGAGCTTTCATGGAGATGGCAGAGCTCATGCTCTCAGAAGGC 240
Db 231 GATTCTTCATCAGTGAAGAGCTTTCATGGAGATGGCAGAGCTCATGCTCTCAGAAGGC 290
Qy 241 TGAAGGATGCAGGTTATGAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCAAGA 300
Db 291 TGAAGGATGCAGGTTATGAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCAAGA 350
Qy 301 GATTCAAGAGCAGACTTCAGGCAGACCTTCAGCGCTTCTCATGGATTCGCCAGCTA 360
Db 351 GATTCAAGAGCAGACTTCAGGCAGACCTTCAGCGCTTCTCATGGATTCGCCAGCTA 410
Qy 361 GCTAATTATCTTCACAGCAAGAGCTGAAGCTAGGATTTATGAGATGTTGGAATAAA 420
Db 411 GCTAATTATCTTCACAGCAAGAGCTGAAGCTAGGATTTATGAGATGTTGGAATAAA 470
Qy 421 ACCTGGCAGGCTTCCTGGAGTTTGGATATACTACAGCAATGATGATGCCAGACCTTTGCT 480

Db 471 ACCTGCCGAGCTTCCTCGGAGTTTGGATACGACATGTATGCCAGACCTTTGCT 530
Qy 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTTGTCTGTGACAGTTTGGAAAATTTG 540
Db 531 GACTGGGAGTAGATCTGCTAAATTTGATGTTTGTCTGTGACAGTTTGGAAAATTTG 590
Qy 541 GCAGATGGTTATAGACATGCTTCCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
Db 591 GCAGATGGTTATAGACATGCTTCCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 650
Qy 601 TCCTGTGAGTGGCTCTTTATATATGTCGCTTCAAAAGCCCAATTATACAGAAATCCGA 660
Db 651 TCCTGTGAGTGGCTCTTTATATATGTCGCTTCAAAAGCCCAATTATACAGAAATCCGA 710
Qy 661 CAGTACTGCAATCACTGCGCAATTTTGTGCAATGATGATTCCTGGAAAAGTATAAAG 720
Db 711 CAGTACTGCAATCACTGCGCAATTTTGTGCAATGATGATTCCTGGAAAAGTATAAAG 770
Qy 721 AGTATCTTGGACTGGAATCTTTTAAACGAGAGAGAAATTTGTGATGTTGCTGACCAAGG 780
Db 771 AGTATCTTGGACTGGAATCTTTTAAACGAGAGAGAAATTTGTGATGTTGCTGACCAAGG 830
Qy 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTTGGCCTCAGCTGGAATCAGCA 840
Db 831 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTTGGCCTCAGCTGGAATCAGCA 890
Qy 841 GTAACCTCAGATGGCCCTCTGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
Db 891 GTAACCTCAGATGGCCCTCTGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 950
Qy 901 CGACACATCAGCCCTCAAGCAAGCTCTCTTCCAGGATAAGGACGTAATTTGCCATCAAT 960
Db 951 CGACACATCAGCCCTCAAGCAAGCTCTCTTCCAGGATAAGGACGTAATTTGCCATCAAT 1010
Qy 961 CAGGACCCCTTGGCAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db 1011 CAGGACCCCTTGGCAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1070
Qy 1021 GAACGACTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCGGAGGAGATGGT 1080
Db 1071 GAACGACTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCGGAGGAGATGGT 1130
Qy 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGGAGTGGCTGTAATCCT 1140
Db 1131 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGGAGTGGCTGTAATCCT 1190
Qy 1141 GCCTGCTTCAATCAGACGCTCCTCCCTGTGAAAAGGAGCTAGGTTCTATGATGACT 1200
Db 1191 GCCTGCTTCAATCAGACGCTCCTCCCTGTGAAAAGGAGCTAGGTTCTATGATGACT 1250
Qy 1201 TCAAGGTTAAGAAAGTACATAAATCCACAGGCACTGTTTGTCTCAGCTATATA 1254
Db 1251 TCAAGGTTAAGAAAGTACATAAATCCACAGGCACTGTTTGTCTCAGCTATATA 1304

RESULT 3

US-10-750-185-40681/c
; Sequence 40681, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482

; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 40681
; LENGTH: 2779
; TYPE: DNA
; ORGANISM: Bovine 19866880781246
US-10-750-185-40681

Query Match 3.0%; Score 37; DB 6; Length 2779;
Best Local Similarity 52.2%; Pred. No. 0.083;
Matches 82; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 631 TTTCAAAGCCCAATTTATACAGAAATCCGACAGTACTGCAATCACTGGCGAAAATTTTGTCT 690
Db 2693 TTTTAGACTCCCTGTTTGTAGTTTGTCTCTATTTATGCGCAGTTCTATCTGTTT 2634
Qy 691 GACATTCATGATTCCTCGGAAAAGTATAAAGAGTATCTTGGACTGGACATCTTTTAACCA 750
Db 2633 GACTCTTGGGGCTCAGGATACACTGTAAAACTTCAGAGGCTTGGTGATCTGTATACCAG 2574
Qy 751 GAGAGAATTTGATGTTGCTGACCAAGGGGTTGGA 787
Db 2573 CAAAGGATATTCAATGTTGCTGTACCAAGTTTGTGA 2537

RESULT 4

US-10-750-185-42255
; Sequence 42255, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 42255
; LENGTH: 1306
; TYPE: DNA
; ORGANISM: Bovine 19866880955804
US-10-750-185-42255

Query Match 2.9%; Score 36.6; DB 6; Length 1306;
Best Local Similarity 53.1%; Pred. No. 0.067;
Matches 78; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 715 ATAAAGAGTATCTTGACTGGACATCTTTTAAACAGAGAGAAATTTGTGATGTTGCTGGA 774
Db 584 ATGCAGAGTTAAATAATAGACAACCTTTAGTCAAGAAGAGCACTTTGTTGTAAGGAATTC 643
Qy 775 CAGGGGGTTGGAATGACCCAGATATGTTAGTGATTTGGCACTTTGGCCTCAGCTGGAAT 834
Db 644 CCAACCCCTGGAACTTTGTGAGTTCTGTAATTTATAGGAGAGCTGGCCTTCCTTGCCTT 703
Qy 835 CAGCAAGTAATCAAGATGGCCCTCTGG 861
Db 704 CTGTCCCTGCCTCAGGTGGTCCCTCG 730

RESULT 5

US-10-995-561-13204/c
; Sequence 13204, Application US/10995561
; Publication No. US20050272054A1

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2006, 00:27:04 ; Search time 410.772 Seconds
(without alignments)
2490.874 Million cell updates/sec

Title: US-10-602-220-29
Perfect score: 89
Sequence: 1 TSRLRSHINPTGTVLLQL 18

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10602220/runat_30122005_140557_5421/app_query.fasta_1.782
-DB=GenEmbl -QWTF=fastcap -SURFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pt0 -NORM=ext -HRAPIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10602220@cgn 1.1.5556 @runat_30122005_140557_5421 -NCPU=6 -ICPU=3
-NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:
1: gb.ba.*
2: gb.in.*
3: gb.env.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pr.*
9: gb.ro.*
10: gb.sts.*
11: gb.sy.*
12: gb.un.*
13: gb.vi.*
14: gb.htg.*
15: gb.pl.*

ALIGNMENTS

RESULT 1
AR101985
LOCUS AR101985 AR101985 210 bp DNA linear PAT 14-FEB-2001
DEFINITION Sequence 19 from patent US 6083725.
ACCESSION AR101985
VERSION AR101985.1 GI:12812783
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 210)
AUTHORS Selden,R.P., Borowski,M., Gillispie,P.P., Kinoshita,C.M.,
Trusco,D.A. and Williams,M.D.
TITLE Transfected human cells expressing human .alpha.-galactosidase A
protein
JOURNAL Patent: US 6083725-A 19 04-JUL-2000;
FEATURES Location/Qualifiers
source 1..210
/organism="unknown"
/mol_type="unassigned DNA"

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	89	100.0	210	6	AR101985 Sequence
2	89	100.0	210	6	BD064230 Therapy f
3	89	100.0	210	6	BD247363 Treatment

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Alignment Scores:
Pred. No.: 3.69e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

```

Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x AR101985 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 155 ACTTCAAGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGCCTTCAGCTA 208

RESULT 2
BD064230
LOCUS BD064230 210 bp DNA linear PAT 27-AUG-2002
DEFINITION Therapy for alpha-galactosidase A deficiency.
ACCESSION BD064230
VERSION BD064230.1 GI:22609833
KEYWORDS JP 2001504324-A/19.
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 210)
AUTHORS Selden,R.F., Borowski,M., Gillespie,P.P., Kinoshita,C.M., Treco,D.A. and Williams,M.D.
TITLE Therapy for alpha-galactosidase A deficiency
JOURNAL Patent: JP 2001504324-A 19 03-APR-2001;
COMMENT TRANSKARYOTIC THERAPIES INC
PN JP 2001504324-A/19
PD 03-APR-2001
PF 12-SEP-1997 JP 1998514004
PR 13-SEP-1996 US 08/712614
PI RICHARD F SELDEN,MARIANNE BOROWSKI,FRANCES
P GILLESPIE,CAROL M
PI KINOSHITA,
PI DOUGLAS A TRECO,MELANIE D WILLIAMS
PC C12N9/40,C12N15/62,C12N15/56,A61K38/47,A61K48/00,C12N5/10 CC
Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
source 1..210
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
ORIGIN
Alignment Scores: 3.69e-07 Length: 210
Pred. No.: 89.00 Matches: 18
Score: 89.00 Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 100.00% Gaps: 0
DB: 6

US-10-602-220-29 (1-18) x BD064230 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 155 ACTTCAAGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGCCTTCAGCTA 208

RESULT 3
BD247363
LOCUS BD247363 210 bp DNA linear PAT 17-JUL-2003
DEFINITION Treatment for alpha-galactosidase A deficiency.
ACCESSION BD247363
VERSION BD247363.1 GI:33057133
KEYWORDS JP 2002538183-A/1.
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 210)
AUTHORS Selden,R.F., Borowski,M., Kinoshita,C.M., Treco,D.A., Williams,M.D., Schuetz,T.J. and Daniel,P.F.
TITLE Treatment for alpha-galactosidase A deficiency
JOURNAL Patent: JP 2002538183-A 1 12-NOV-2002;

TRANSKARYOTIC THERAPIES INC
OS Artificial Sequence
PN JP 2002538183-A/1
PD 12-NOV-2002
PF 09-MAR-2000 JP 2000603353
PR 11-MAR-1999 US 09/266014
PI RICHARD F SELDEN,MARIANNE BOROWSKI,CAROL
M KINOSHITA,DOUGLAS A
PI TRECO,
PI MELANIE D WILLIAMS,THOMAS J SCHUETZ,PETER F DANIEL,PC
A61K38/43,A61P3/00,A61P3/08,C12N9/24//C12N15/09,A61K37/48, PC
C12N15/00
CC Description of Artificial Sequence: Human fibroblast library
CC probe: exon
CC 7, including amplification primers.
FH Key Location/Qualifiers
FT source 1..210
/organism='Artificial Sequence'.
FEATURES
source 1..210
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
ORIGIN
Alignment Scores: 3.69e-07 Length: 210
Pred. No.: 89.00 Matches: 18
Score: 89.00 Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 100.00% Gaps: 0
DB: 6

US-10-602-220-29 (1-18) x BD247363 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 155 ACTTCAAGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGCCTTCAGCTA 208

RESULT 4
CS105028
LOCUS CS105028 210 bp DNA linear PAT 10-JUN-2005
DEFINITION Sequence 19 from Patent EP1538202.
ACCESSION CS105028
VERSION CS105028.1 GI:67512960
KEYWORDS unidentified
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1
AUTHORS Selden,R.F., Borowski,M., Treco,D.A., Gillespie,P.P., Kinoshita,C.M. and Williams,M.D.
TITLE Production of human alpha-galactosidase A
JOURNAL Patent: EP 1538202-A 19 08-JUN-2005;
Transkaryotic Therapies, Inc. (US)
FEATURES
source 1..210
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
ORIGIN
Alignment Scores: 3.69e-07 Length: 210
Pred. No.: 89.00 Matches: 18
Score: 89.00 Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 100.00% Gaps: 0
DB: 6

US-10-602-220-29 (1-18) x CS105028 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 155 ACTTCAAGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGCCTTCAGCTA 208

```

Db 155 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTCTTCAGCTA 208
|||||
RESULT 5
AR234437 AR234437 210 bp DNA linear PAT 20-DEC-2002
LOCUS
DEFINITION Sequence 1 from patent US 6458574.
ACCESSION AR234437
VERSION AR234437.1 GI:27277133
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 210)
AUTHORS Selden,R.F., Borowski,M., Kinoshita,C.M., Treco,D.A.,
Williams,M.D., Schuetz,T.J. and Daniel,P.F.
TITLE Treatment of a .alpha.-galactosidase a deficiency
JOURNAL Patent: US 6458574-A 1 01-OCT-2002;
Transkaryotic Therapies, Inc.; Cambridge, MA
FEATURES
source
1..210
Location/Qualifiers
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 3.69e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-602-220-29 (1-18) x AR234437 (1-210)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
|||||
Db 155 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTCTTCAGCTA 208
|||||
RESULT 6
AR232463 AR232463 210 bp DNA linear PAT 17-AUG-2003
LOCUS
DEFINITION Sequence 19 from patent US 6566099.
ACCESSION AR322463
VERSION AR322463.1 GI:33708228
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 210)
AUTHORS Selden,R.F., Borowski,M., Gillispie,F.P., Kinoshita,C.M.,
Treco,D.A. and Williams,M.D.
TITLE Nucleic acid encoding a chimeric polypeptide
JOURNAL Patent: US 6566099-A 19 20-MAY-2003;
Transkaryotic Therapies, Inc.; Cambridge, MA
FEATURES
source
1..210
Location/Qualifiers
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 3.69e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-602-220-29 (1-18) x AR322463 (1-210)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
|||||

Db 155 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTCTTCAGCTA 208
|||||
RESULT 7
AR372139 AR372139 210 bp DNA linear PAT 12-SEP-2003
LOCUS
DEFINITION Sequence 19 from patent US 6395884.
ACCESSION AR372139
VERSION AR372139.1 GI:34609448
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 210)
AUTHORS Selden,R.F., Borowski,M., Gillispie,F.P., Kinoshita,C.M.,
Treco,D.A. and Williams,M.D.
TITLE Therapy for .alpha.-galactosidase a deficiency
JOURNAL Patent: US 6395884-A 19 28-MAY-2002;
Transkaryotic Therapies, Inc.; DE
FEATURES
source
1..210
Location/Qualifiers
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 3.69e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-602-220-29 (1-18) x AR372139 (1-210)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
|||||
Db 155 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTCTTCAGCTA 208
|||||
RESULT 8
CQ684109 CQ684109 227 bp DNA linear PAT 03-FEB-2004
LOCUS
DEFINITION Sequence 29035 from Patent WO202070737.
ACCESSION CQ684109
VERSION CQ684109.1 GI:42208348
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Liew,C.C., Marshall,W.E. and Zhang,H.
TITLE Compositions and methods relating to osteoarthritis
JOURNAL Patent: WO 02070737-A 29035 12-SEP-2002;
Chondrogene Inc. (CA)
FEATURES
source
1..227
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 4.01e-07 Length: 227
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-602-220-29 (1-18) x CQ684109 (1-227)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
|||||

Db 144 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTA 197
|||||
RESULT 9
LOCUS CQ685605 270 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 30531 from Patent WO02070737.
ACCESSION CQ685605
VERSION CQ685605.1 GI:42213120
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
AUTHORS Liew, C.C., Marshall, W.E. and Zhang, H.
TITLE Compositions and methods relating to osteoarthritis
JOURNAL Patent: WO 02070737-A 30531 12-SEP-2002;
Chondrogene Inc. (CA)
FEATURES
source location/Qualifiers
1..270
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 4.82e-07 Length: 270
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-602-220-29 (1-18) x CQ685605 (1-270)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 174 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTA 227
|||||
RESULT 10
LOCUS I30314 297 bp DNA linear PAT 06-FEB-1997
DEFINITION Sequence 11 from patent US 5580757.
ACCESSION I30314
VERSION I30314.1 GI:1821105
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
AUTHORS Desnick, R.J., Bishop, D.F. and Ioannou, Y.A.
TITLE Cloning and expression of biologically active .alpha.-galactosidase
A as a fusion protein
JOURNAL Patent: US 5580757-A 11 03-DEC-1996;
FEATURES
source location/Qualifiers
1..297
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Alignment Scores:
Pred. No.: 5.34e-07 Length: 297
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-602-220-29 (1-18) x I30314 (1-297)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18

Db 7 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTA 60
|||||
RESULT 11
LOCUS AX030666 297 bp DNA linear PAT 20-SEP-2000
DEFINITION Sequence 8 from Patent EP1020528.
ACCESSION AX030666
VERSION AX030666.1 GI:10278187
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
unclassified sequences.
REFERENCE
AUTHORS Ioannou, Y.A., Bishop, D.F. and Deanick, R.J.
TITLE Method for producing secreted proteins
JOURNAL Patent: EP 1020528-A 8 19-JUL-2000;
SINAI SCHOOL MEDICINE (US)
FEATURES
source location/Qualifiers
1..297
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"
<1..279
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC09606.1"
/db_xref="GI:10278188"
/translation="EWTSRLRSHINPTGTVLLQLENTWQMSLKDLLPAGPAQHDEAAQQ
NAFYQVLNMPNLNADQRNGFTQSLKDDPSQSANVLGEAQKLNDSQAPK"
ORIGIN
Alignment Scores:
Pred. No.: 5.34e-07 Length: 297
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-602-220-29 (1-18) x AX030666 (1-297)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 7 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTA 60
|||||
RESULT 12
LOCUS ARI43996 1158 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 53 from patent US 6210666.
ACCESSION ARI43996
VERSION ARI43996.1 GI:15105863
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
AUTHORS Miyamura, N.
TITLE Truncated .alpha.-galactosidase A to treat fabry disease
JOURNAL Patent: US 6210666-A 53 03-APR-2001;
FEATURES
source location/Qualifiers
1..1158
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Alignment Scores:
Pred. No.: 2.28e-06 Length: 1158
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x ARI43996 (1-1158)

QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACGTGTTTGCCTTCAGCTA 1158

RESULT 13
LOCUS ARI43995 1161 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 52 from patent US 6210666.
ACCESSION ARI43995
VERSION ARI43995.1 GI:15105862
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1161)
AUTHORS Miyamura,N.
TITLE Truncated .alpha.-galactosidase A to treat fabry disease
JOURNAL Patent: US 6210666-A 52 03-APR-2001;
FEATURES Location/Qualifiers
source 1..1161
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:
Pred. No.: 2.28e-06 Length: 1161
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x ARI43995 (1-1161)

QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACGTGTTTGCCTTCAGCTA 1158

RESULT 14
LOCUS ARI43994 1164 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 51 from patent US 6210666.
ACCESSION ARI43994
VERSION ARI43994.1 GI:15105861
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1164)
AUTHORS Miyamura,N.
TITLE Truncated .alpha.-galactosidase A to treat fabry disease
JOURNAL Patent: US 6210666-A 51 03-APR-2001;
FEATURES Location/Qualifiers
source 1..1164
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:
Pred. No.: 2.29e-06 Length: 1164
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x ARI43994 (1-1164)

QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||

Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACGTGTTTGCCTTCAGCTA 1158

RESULT 15

LOCUS ARI43993 1167 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 50 from patent US 6210666.
ACCESSION ARI43993
VERSION ARI43993.1 GI:15105860
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1167)
AUTHORS Miyamura,N.
TITLE Truncated .alpha.-galactosidase A to treat fabry disease
JOURNAL Patent: US 6210666-A 50 03-APR-2001;
FEATURES Location/Qualifiers
source 1..1167
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:
Pred. No.: 2.29e-06 Length: 1167
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x ARI43993 (1-1167)

QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACGTGTTTGCCTTCAGCTA 1158

US-10-602-220-29 (1-18) x ARI43993 (1-1167)

QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||

Search completed: January 1, 2006, 03:35:28
Job time : 412.772 secs

This Page Blank (uspto)

November 2005

Published_Applications Nucleic Acid and Published_Applications Amino Acid database searches now generate two sets of results each. The Published_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published_Applications_New databases; older published applications make up the Published_Applications_Main databases.

Searches run against Nucleic Acid Published_Applications produce two sets of results, with the extensions **.rnpbm** (Published_Applications_NA_Main) and **.rnpbn** (Published_Applications_NA_New).

Searches run against Amino Acid Published_Applications produce two sets of results, with the extensions **.rapbm** (Published_Applications_AA_Main) and **.rapbn** (Published_Applications_AA_New).

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 31, 2005, 23:48:43 ; Search time 48.2483 Seconds
(without alignments)
2486.399 Million cell updates/sec

Title: US-10-602-220-29

Perfect score: 89
Sequence: 1 TSLRLSHNPTGTVLLQL 18

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10602220/runat_30122005_140557_5411/app_query.fasta.1.782
-DB=N_Geneseq -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODES=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10602220@cgn 1.1.900 @runat_30122005_140557_5411 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=110 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq 21.*

- 1: Geneseq1980s.*
- 2: Geneseq1990s.*
- 3: Geneseq2000s.*
- 4: Geneseq2001as.*
- 5: Geneseq2001bs.*
- 6: Geneseq2002as.*
- 7: Geneseq2002bs.*
- 8: Geneseq2003as.*
- 9: Geneseq2003bs.*
- 10: Geneseq2003cs.*
- 11: Geneseq2003ds.*
- 12: Geneseq2004as.*
- 13: Geneseq2004bs.*
- 14: Geneseq2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	100.0	210	2	Aav31800 Nucleotid
2	89	100.0	210	3	Aaa75432 Nucleotid
3	89	100.0	210	3	Aaa70681 Human alp
4	89	100.0	210	6	Aad42074 Probe use

5	89	100.0	210	9	ACA62301	Ac62301 Probe for
6	89	100.0	210	10	AAD60098	Aad60098 210 bp pr
7	89	100.0	210	10	ADH00941	Adh00941 Probe fro
8	89	100.0	210	10	ADH54465	Adh54465 Probe to
9	89	100.0	210	12	ADG47825	Adg47825 Probe use
10	89	100.0	297	2	AAQ66242	Aaq66242 Protein A
11	89	100.0	1158	4	AAC85748	Aac85748 Mutant de
12	89	100.0	1161	4	AAC85747	Aac85747 Mutant de
13	89	100.0	1164	4	AAC85746	Aac85746 Mutant de
14	89	100.0	1167	4	AAC85745	Aac85745 Mutant de
15	89	100.0	1170	4	AAC85744	Aac85744 Mutant de
16	89	100.0	1173	4	AAC85743	Aac85743 Mutant de
17	89	100.0	1176	4	AAC85742	Aac85742 Mutant de
18	89	100.0	1179	4	AAC85741	Aac85741 Mutant de
19	89	100.0	1182	4	AAC85740	Aac85740 Mutant de
20	89	100.0	1188	4	AAC85739	Aac85739 Mutant de
21	89	100.0	1194	4	AAC85738	Aac85738 Alpha-gal
22	89	100.0	1197	3	AA707000	Aa707000 Human alp
23	89	100.0	1197	6	AAD42079	Aad42079 Human alp
24	89	100.0	1197	9	ACA62304	Ac62304 cDNA sequ
25	89	100.0	1197	10	AAD60103	Aad60103 Human alp
26	89	100.0	1197	10	ADH00921	Adh00921 Human alp
27	89	100.0	1197	10	ADH54443	Adh54443 Human alp
28	89	100.0	1197	12	ADG47805	Adg47805 Human alp
29	89	100.0	1214	11	ADI31675	Adi31675 Human cDN
30	89	100.0	1214	13	ADS83742	Ad83742 Human lym
31	89	100.0	1254	6	AAD45225	Aad45225 Human rga
32	89	100.0	1254	10	ADD84754	Adm84754 Human alp
33	89	100.0	1254	12	ADM48684	Adm48684 Human wil
34	89	100.0	1254	13	ADU66919	Adu66919 Human alp
35	89	100.0	1254	14	AEA27448	Aea27448 Human alp
36	89	100.0	1266	6	AAD45223	Aad45223 Human rga
37	89	100.0	1266	10	ADD84750	Adm84750 Human alp
38	89	100.0	1266	12	ADJ88276	Adj88276 Human WT
39	89	100.0	1266	12	ADM48680	Adm48680 Human wil
40	89	100.0	1266	13	ADU66915	Adu66915 Human alp
41	89	100.0	1266	14	AEA27444	Aea27444 Human alp
42	89	100.0	1272	6	AAD45226	Aad45226 Human rga
43	89	100.0	1272	10	ADD84756	Adm84756 Human alp
44	89	100.0	1272	12	ADJ88282	Adj88282 Human WT
45	89	100.0	1272	12	ADM48686	Adm48686 Human wil

ALIGNMENTS

RESULT 1
AAV31800
ID AAV31800 standard; DNA; 210 BP.
XX
AC AAV31800;
XX
DT 21-AUG-1998 (first entry)
XX
DE Nucleotide sequence of the human alpha-gal A gene probe.
XX

Probe; hybridisation; human alpha-galactosidase A gene; alpha-gal A gene;
glycosylated human alpha-gal A; Fabry disease; ss;
alpha-gal A deficiency; Fabry; X-linked; lysosomal storage disease.

OS Synthetic.

OS Homo sapiens.

XX

PN WO9811206-A2.

XX

PD 19-MAR-1998.

XX

PF 12-SEP-1997; 97WO-US016603.

XX

PR 13-SEP-1996; 96US-00712614.

XX

PA (TRAN-) TRANSKARYOTIC THERAPIES INC.

XX

PI Seiden RF, Borowski M, Gillespie FP, Kinoshita CM, Treco DA;

PI Williams MD;
 XX WPI; 1998-207375/18.
 XX
 XX Treatment of alpha-galactosidase A deficiency in, e.g. Fabry disease -
 PT comprises use of genetically modified human cells that express alpha-
 PT galactosidase.
 XX
 XX Disclosure; Fig 1; 78pp; English.
 XX
 XX This is the nucleotide sequence of the probe used in the to isolate the
 CC human alpha-galactosidase A gene (alpha-gal A), used in the method of the
 CC invention, which involves the production of genetically modified human
 CC cells that express alpha-galactosidase. The cultured human cells can be
 CC used for the production of glycosylated human alpha-gal A and can be used
 CC for the treatment of patients with an alpha-gal A deficiency, e.g. Fabry
 CC disease (an X-linked inherited lysosomal storage disease)
 XX
 XX SQ Sequence 210 BP; 52 A; 51 C; 52 G; 55 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 8.41e-07 Length: 210
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0
 US-10-602-220-29 (1-18) x AAV31800 (1-210)
 QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 DB 155 ACTTCAAGGTTAAGAAGTCACATAAAATCCACAGGCACTGTTTTCAGCTA 208
 RESULT 2
 AAA75432
 ID AAA75432 standard; DNA; 210 BP.
 XX
 XX AC AAA75432;
 XX
 XX 15-JAN-2001 (first entry)
 DT
 XX Nucleotide sequence of a probe from exon 7 of human alpha-Gal A gene.
 DE
 XX Human; alpha-galactosidase A; alpha-Gal A; lysosomal glycohydrolase;
 XX Fabry's disease; cardiovascular abnormality; probe;
 KW left ventricular hypertrophy; LVH; ss.
 XX
 XX Synthetic.
 OS
 XX Homo sapiens.
 OS
 XX WO200053730-A2.
 FN
 XX 14-SEP-2000.
 PD
 XX
 XX 09-MAR-2000; 2000WO-US006118.
 PF
 XX
 XX 11-MAR-1999; 99US-00266014.
 PR
 XX (TRAN-) TRANSKARYOTIC THERAPIES INC.
 PA
 XX Selden RF, Borowski M, Kinoshita CM, Treco DA, Williams MD;
 PI Schuetz TJ, Daniel PF;
 PI
 XX WPI; 2000-594316/56.
 DR
 XX Pure human alpha-Gal A preparation for treating alpha-galactosidase A
 PT deficiency associated disorders e.g. Fabry's disease, cardiovascular
 PT abnormality, has various alpha-Gal A glycoforms and is free of lectins.
 XX
 XX Example 1; Fig 1; 92pp; English.
 PS
 XX The present sequence represents a probe derived from exon 7 of the human

CC alpha-galactosidase A (alpha-Gal A) gene. Alpha-Gal A is lysosomal
 CC glycohydrolase which cleaves the terminal alpha-galactosyl moieties of
 CC various glycoconjugates. The protein is useful for preparing a medicament
 CC for treating alpha-gal A deficiency associated disorders, e.g. typical
 CC variant of Fabry's disease, cardiovascular abnormality, preferably left
 CC ventricular hypertrophy (LVH). Alpha-Gal A produced by the invention has
 CC an extended circulating half-life and increased uptake in specific
 CC tissues other than live, compared with alpha-Gal A produced by other
 CC methods
 XX
 XX SQ Sequence 210 BP; 52 A; 51 C; 52 G; 55 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 8.41e-07 Length: 210
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 3 Gaps: 0
 US-10-602-220-29 (1-18) x AAA75432 (1-210)
 QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 DB 155 ACTTCAAGGTTAAGAAGTCACATAAAATCCACAGGCACTGTTTTCAGCTA 208
 RESULT 3
 AAA70681
 ID AAA70681 standard; DNA; 210 BP.
 XX
 XX AC AAA70681;
 XX
 XX 31-JAN-2001 (first entry)
 DT
 XX Human alpha-Gal A gene exon 7 probe.
 DE
 XX Nephrotropic; gene therapy; alpha-galactosidase A deficiency; ss;
 KW Fabry disease; secretion; human growth hormone; hGH; signal peptide;
 KW enzyme replacement therapy; probe hybridisation.
 XX
 XX Homo sapiens.
 OS
 XX US6083725-A.
 PN
 XX 04-JUL-2000.
 PD
 XX 12-SEP-1997; 97US-00928881.
 PF
 XX 13-SEP-1996; 96US-0026041P.
 PR
 XX (TRAN-) TRANSKARYOTIC THERAPIES INC.
 PA
 XX Kinoshita CM, Treco DA, Williams MD, Selden RF, Borowski M;
 XX Gillispie FP;
 PI
 XX WPI; 2000-464341/40.
 DR
 XX New DNA comprising a human growth hormone signal peptide and human alpha-
 PT galactosidase A polypeptide and cells expressing the DNA, useful for
 PT treating Fabry disease and in vitro protein production for enzyme
 PT replacement therapy.
 XX
 XX Example 1; Fig 1; 32pp; English.
 PS
 XX The invention relates to the treatment of a patient having an alpha-
 CC galactosidase A (aga) deficiency, e.g. Fabry disease, by using cells
 CC overexpressing and secreting aga or purified aga. This sequence
 CC represents a 270 bp fragment of the human aga exon 7 sequence for use as
 CC a probe to isolate the complete aga gene sequence (AAA70685). The probe
 CC was amplified by the primers AAA70679-A70680. For expression and
 CC secretion of the aga, the protein is fused to a signal peptide especially
 CC the human growth hormone (hGH) signal peptide. Primary, secondary, or
 CC immortalised human cells genetically modified to overexpress human aga

CC are also useful for in vitro protein production, to produce protein which
 CC may be purified for enzyme replacement therapy. The method allows the
 CC production of large quantities of appropriately glycosylated and
 CC therapeutically useful human alpha

SQ Sequence 210 BP; 52 A; 51 C; 52 G; 55 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8.41e-07 Length: 210
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x AAA70681 (1-210)

QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 |||||
 DB 155 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCCTTCAGCTA 208

RESULT 4

AAD42074

ID AAD42074 standard; DNA; 210 BP.

XX

AC AAD42074;

XX

DT 04-NOV-2002 (first entry)

XX

DE Probe used to isolate human alpha-gal A cDNA.

XX

KW Human; alpha-galactosidase A; alpha-gal A deficiency; Fabry disease;
 KW enzyme; probe; ss.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT primer_bind complement(1..21)

FT /*tag= a

FT /bound_moiety= "Primer 1"

FT primer_bind 190..210

FT /*tag= b

FT /bound_moiety= "Primer 2"

FT

XX

PN US6395884-B1.

XX

PD 28-MAY-2002.

XX

PF 06-APR-2000; 2000US-00543921.

XX

PR 13-SEP-1996; 96US-0026041P.

XX

PR 12-SEP-1997; 97US-00928881.

XX

PA (TRAN-) TRANSKARYOTIC THERAPIES INC.

XX

PI Seiden RF, Borowski M, Gillispie FP, Kinoshita CM, Treco DA;
 PI Williams MD;

XX

DR WPI; 2002-526374/56.

XX

PT Purifying human alpha-galactosidase A from a cell for use in treating
 PT Fabry disease, by passing sample containing the enzyme over hydrophobic
 PT interaction resin, heparin resin, hydroxyapatite and anion exchange
 PT resin.

XX

PS Example 1; Col 31-32; 31pp; English.

XX

CC The invention relates to a method of purifying human alpha-galactosidase
 CC A (alpha-gal A) from a human cell. The method involves a chromatography
 CC step which comprises passing human alpha-gal A containing sample over a
 CC hydrophobic interaction resin. The method is useful for purifying alpha-
 CC gal A from a human cell, for use in treating an individual suspected of
 CC having alpha-gal A deficiency such as Fabry disease. The present sequence

CC is a probe used to isolate human alpha-gal A cDNA

SQ Sequence 210 BP; 52 A; 51 C; 52 G; 55 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8.41e-07 Length: 210
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x AAD42074 (1-210)

QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 |||||
 DB 155 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCCTTCAGCTA 208

RESULT 5

ACA62301

ID ACA62301 standard; DNA; 210 BP.

XX

AC ACA62301;

XX

DT 13-AUG-2003 (first entry)

XX

DE Probe for cDNA encoding human alpha-galactosidase A glycoprotein.

XX

KW Human; alpha-galactosidase A glycoprotein; alpha-gal A; glycan;
 KW complex-type glycan; alpha-galactosidase A deficiency; Fabry's disease;
 KW cardiovascular abnormality; ventricular enlargement;
 KW left ventricular hypertrophy; mitral valve insufficiency;
 KW renal abnormality; sialylation; nephrotropic; probe; ss.

XX

OS Homo sapiens.

XX

PN US6458574-B1.

XX

PD 01-OCT-2002.

XX

PF 11-MAR-1999; 99US-00266014.

XX

PR 13-SEP-1996; 96US-0026041P.

XX

PR 12-SEP-1997; 97US-00928881.

XX

PR 12-SEP-1997; 97WO-US016603.

XX

PA (TRAN-) TRANSKARYOTIC THERAPIES INC.

XX

PI Seiden RF, Borowski M, Kinoshita CM, Treco DA, Williams MD;
 PI Schuetz TJ, Daniel PF;

XX

DR WPI; 2003-465541/44.

XX

PT Human alpha-galactosidase A glycoprotein preparation containing complex-
 PT type glycans, useful for treating alpha-galactosidase A deficiency,
 PT especially Fabry's disease.

XX

PS Example 1; Fig 1; 39pp; English.

XX

CC The present invention relates to the isolation of highly purified human
 CC alpha-galactosidase A (alpha-gal A) glycoprotein, and preparations
 CC containing alpha-gal A, where greater than 50% of the total glycans of
 CC the preparation are complex-type glycans, produced by providing a human
 CC cell genetically modified to express alpha-gal A polypeptide, and
 CC purifying the polypeptide from the human cell or its culture medium. The
 CC alpha-gal A preparations are useful for the treatment of alpha-
 CC galactosidase A deficiency, especially Fabry's disease e.g. specific
 CC populations of Fabry patients with predominantly cardiovascular
 CC abnormalities such as ventricular enlargement (e.g. left ventricular
 CC hypertrophy and/or mitral valve insufficiency), or Fabry patients with
 CC predominantly renal involvement. The alpha-gal A preparations have
 CC extended circulating half-life in mammalian hosts, as improved
 CC sialylation of alpha-gal A preparations enhances the circulatory half-

CC life of exogenous alpha-Gal A. The present sequence represents a probe
CC used to isolate cDNA encoding human alpha-Gal A in the examples of the
CC present invention

XX SQ Sequence 210 BP; 52 A; 51 C; 52 G; 55 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 8.41e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-10-602-220-29 (1-18) x ACA62301 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 155 ACTTCAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCAGCTA 208

RESULT 6

AAAD60098
ID AAD60098 standard; DNA; 210 BP.

XX AC AAD60098;

XX DT 18-DEC-2003 (first entry)

XX DE 210 bp probe used to isolate human alpha-gal A gene.

XX KW Human growth hormone; glucocerebrosidase; gene therapy; cardiovascular;
XX KW Gaucher's disease; Fabry disease; alpha-galactosidase A deficiency; hGH;
XX KW CNS; nephrotropic; human; probe; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
XX FT primer_bind complement(1..22)
FT /*tag= a
FT primer_bind /bound_moiety= "primer 1"
FT 190..210
FT /*tag= b
FT /bound_moiety= "primer 2"

XX US6566099-B1.

XX PD 20-MAY-2003.

XX PF 27-JAN-2000; 2000US-00491759.

XX PR 13-SEP-1996; 96US-0026041P.

XX PR 12-SEP-1997; 97US-00928881.

XX (TRAN-) TRANSKARYOTIC THERAPIES INC.

XX Seiden RF, Borowski M, Gillispie FP, Kinoshita CM, Treco DA;
XX Williams MD;

XX WPI; 2003-719643/68.

XX Novel polypeptide comprising signal peptide of human growth hormone
XX linked to glucocerebrosidase useful for treating disorders associated
XX with enzyme deficiency in sphingolipid degradation such as gaucher
XX disease.

XX Example 1; Fig 1; 30pp; English.

XX The present invention relates to a novel polypeptide comprising signal
XX peptide of human growth hormone (hGH) linked to glucocerebrosidase. The
XX invention is useful for treating disorders associated with alpha-
XX galactosidase A deficiency and sphingolipid degradation such as Gaucher's
XX disease and Fabry disease. The invention is also useful in gene therapy.
XX The present sequence is a 210 bp probe used to isolate human alpha-Gal A

CC gene

XX SQ Sequence 210 BP; 52 A; 51 C; 52 G; 55 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8.41e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-602-220-29 (1-18) x AAD60098 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 155 ACTTCAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCAGCTA 208

RESULT 7

ADH00941
ID ADH00941 standard; DNA; 210 BP.

XX AC ADH00941;

XX DT 11-MAR-2004 (first entry)

XX DE Probe from human alpha-galactosidase A (alpha-Gal A) gene.

XX KW Human; ss; alpha-galactosidase A; alpha-Gal A; enzyme therapy; therapy;
XX KW Fabry disease; cardiovascular abnormality; left ventricular hypertrophy;
XX KW cardiand; probe.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
XX FT primer_bind 1..22
FT /*tag= a
FT primer_bind 190..210
FT /*tag= b

XX US2003077806-A1.

XX PD 24-APR-2003.

XX PF 07-JUN-2002; 2002US-00165060.

XX PR 13-SEP-1996; 96US-0026041P.

XX PR 12-SEP-1997; 97US-00928881.

XX PR 12-SEP-1997; 97WO-US016603.

XX PR 11-MAR-1999; 99US-00266014.

XX (TRAN-) TRANSKARYOTIC THERAPIES INC.

XX Seiden RF, Borowski M, Kinoshita CM, Treco DA, Williams MD;
XX Schuetz TJ, Daniel PF;

XX WPI; 2003-567381/53.

XX Composition useful for treating Fabry disease or atypical variant of
XX Fabry disease comprising a human alpha-galactosidase A preparation,
XX purified to at least 98% homogeneity.

XX Example 1; SEQ ID NO 1; 42pp; English.

XX The invention relates to a composition comprising a human alpha-
XX galactosidase A (alpha-Gal A) preparation, purified to at least 98%
XX homogeneity, as measured by sodium dodecyl sulphate polyacrylamide gel
XX electrophoresis (SDS-PAGE) or reverse phase HPLC, where the human alpha-
XX Gal A preparation has a specific activity of at least 2.0x10⁶ units/mg
XX protein. The invention is useful in enzyme therapy. The
XX invention is useful for treating a patient with Fabry disease, or for
XX treating a patient with an atypical variant of Fabry disease (e.g.
XX cardiovascular abnormality such as left ventricular hypertrophy), where

CC the alpha-Gal A preparation is administered to the patient in a dose
 CC between 0.05-5 mg of alpha-Gal A preparation per kg body weight weekly or
 CC biweekly. The present sequence is a probe from human alpha-Gal A gene
 CC used to isolate the cDNA.

SQ Sequence 210 BP; 52 A; 51 C; 52 G; 55 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8.41e-07 Length: 210
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-602-220-29 (1-18) x ADH00941 (1-210)

QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 DB 155 ACTTCAGGTTAAGAAGTCACATAATCCACAGGCACCTGTTTTCCTCAGCTA 208

RESULT 8

ADH54465
 ID ADH54465 standard; cDNA; 210 BP.

XX AC ADH54465;

XX DT 25-MAR-2004 (first entry)

XX DE Probe to isolate human alpha-galactosidase A (alpha-gal A) cDNA.

XX KW Alpha-galactosidase A; nephrotropic; nephrotropic; gene therapy;

XX KW Fabry disease; human; ss; probe.

XX OS Homo sapiens.

XX PN US2003152560-A1.

XX PD 14-AUG-2003.

XX PF 12-DEC-2002; 2002US-00318905.

XX PR 13-SEP-1996; 96US-0026041P.

XX PR 12-SEP-1997; 97US-00928881.

XX PR 27-JAN-2000; 2000US-00491759.

XX PA (TRAN-) TRANSKARYOTIC THERAPIES INC.

XX PI Selden RF, Borowski M, Gillispie FP, Kinoshita CM, Treco DA;

XX PI Williams MD;

XX DR WPI; 2003-897682/82.

XX Therapeutic method in alpha-galactosidase A deficiency using genetically
 PT modified cells overexpressing and secreting the human alpha-galactosidase
 PT A, useful particularly for treating Fabry disease.

PS Example 1; SEQ ID NO 19; 33pp; English.

XX The invention relates to a method of treatment which comprises
 CC identifying a patient suspected of having an alpha-galactosidase A
 CC deficiency, and introducing into the patient a human cell genetically
 CC modified to overexpress and secrete human alpha-gal A. The invention is
 CC useful in gene therapy. The methods and compositions of the present
 CC invention are useful for treating disorders with alpha-galactosidase
 CC deficiency, such as Fabry disease. The present sequence is a probe used
 CC to isolate alpha-galactosidase A cDNA.

SQ Sequence 210 BP; 52 A; 51 C; 52 G; 55 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8.41e-07 Length: 210
 Score: 89.00 Matches: 18

Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-602-220-29 (1-18) x ADH54465 (1-210)

QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 DB 155 ACTTCAGGTTAAGAAGTCACATAATCCACAGGCACCTGTTTTCCTCAGCTA 208

RESULT 9

ADG47825

ID ADG47825 standard; DNA; 210 BP.

XX AC ADG47825;

XX DT 11-MAR-2004 (first entry)

XX DE Probe used to isolate human alpha-Gal A cDNA.

XX KW Alpha-galactosidase A; alpha-Gal A; Fabry disease;
 KW cardiovascular abnormality; ventricular enlargement;
 KW left ventricular hypertrophy; LVH; mitral valve insufficiency;
 KW renal involvement; enzyme therapy; human; probe; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 FT primer_bind complement(1..22)
 FT primer_bind /*tag= a
 FT primer_bind 190..210
 FT primer_bind /*tag= b

XX PN US2003113894-A1.

XX PD 19-JUN-2003.

XX PF 10-JUN-2002; 2002US-00165968.

XX PR 13-SEP-1996; 96US-0026041P.

XX PR 12-SEP-1997; 97US-00928881.

XX PR 12-SEP-1997; 97WO-US016603.

XX PR 11-MAR-1999; 99US-00266014.

XX PA (TRAN-) TRANSKARYOTIC THERAPIES INC.

XX PI Selden RF, Borowski M, Kinoshita CM, Treco DA, Williams MD;

XX PI Schuetz TJ, Daniel PF;

XX DR WPI; 2004-009047/01.

XX Composition useful for treating Fabry disease or atypical variant of
 PT Fabry disease comprising a human alpha-galactosidase A preparation,
 PT purified to at least 98 percent homogeneity.

PS Example 1; SEQ ID NO 1; 43pp; English.

XX The present invention provides highly purified alpha-galactosidase A
 CC (alpha-Gal A) and various methods for purifying it; alpha-Gal A
 CC preparations with altered charge and methods for making those
 CC preparations; alpha-Gal A preparations that have an extended circulating
 CC half-life in a mammalian host and methods for making the same; and
 CC methods and dosages for administering an alpha-Gal preparation to a
 CC subject. The alpha-Gal A preparation of the invention is useful for
 CC creating a patient with Fabry disease, or atypical variants of Fabry
 CC disease e.g. cardiovascular abnormality such as ventricular enlargement
 CC e.g. left ventricular hypertrophy (LVH) and/or mitral valve insufficiency
 CC or Fabry patients with predominantly renal involvement. The present
 CC sequence is a probe used to isolate human alpha-Gal A cDNA.

SQ Sequence 210 BP; 52 A; 51 C; 52 G; 55 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 8.41e-07 Length: 210
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 12 Gaps: 0

US-10-602-220-29 (1-18) x ADG47825 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 |||||
 Db 155 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCAGCTA 208
 |||||

RESULT 10
 AAQ66242
 ID AAQ66242 standard; DNA; 297 BP.
 XX
 AC AAQ66242;
 XX
 DT 25-MAR-2003 (revised)
 DT 15-NOV-1994 (first entry)
 XX
 DE Protein A domain E.
 XX
 KW Alpha-galactosidase A; Fabry disease; blood group O; blood group B;
 KW COS-1; enzyme replacement therapy; glycoconjugate; protein A;
 KW IgG binding domain; fusion protein; alpha-Gal A; ds.
 XX
 OS Staphylococcus aureus.
 XX
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 1..279
 FT /*tag= a
 XX
 XX
 XX WO9412628-A1.
 XX
 XX 09-JUN-1994.
 XX
 XX 30-NOV-1993; 93WO-US011539.
 XX
 XX 30-NOV-1992; 92US-00983451.
 XX
 XX (MOUN) MOUNT SINAI SCHOOL MEDICINE.
 XX
 XX Deanick RJ, Bishop DF, Ioannou YA;
 XX
 XX WPI; 1994-200257/24.
 XX
 XX P-PSDB; AAR53771.
 XX
 XX Prodn of human alpha-galactosidase A - by culturing cells contg the
 PT coding sequence and the beta-galactosidase alpha-2,6-sialyl transferase
 PT gene and regulation sequences.
 XX
 XX Disclosure; Page 107-108; 156pp; English.

XX The cDNA sequence (AAQ66241) of human alpha-galactosidase A (alpha-Gal A)
 CC was determined. A fusion protein of alpha-Gal A cDNA and protein A IGG
 CC binding domain E (AAQ66242, AAR53771) was expressed in COS-1 cells. The
 CC alpha-Gal A cDNA was amplified by PCR using primers (AAQ66243-44)
 CC designed to delete the termination codon, to add a 3' collagenase
 CC cleavage consensus sequence and to include a 5' EcoRI recognition site.
 CC The protein A domain E was amplified with the collagenase consensus
 CC sequence in the 5' oligonucleotide using the primers given in AAQ66245-
 CC 46. The enzyme can be used for Fabry disease enzyme replacement therapy,
 CC to convert blood group B to O, or to hydrolyze alpha-D-galactosyl
 CC residues from glycoconjugates. (Updated on 25-MAR-2003 to correct EN
 CC field.)
 XX

SQ Sequence 297 BP; 103 A; 64 C; 54 G; 76 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1.29e-06 Length: 297

Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-10-602-220-29 (1-18) x AAQ66242 (1-297)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 |||||
 Db 7 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCAGCTA 60
 |||||

RESULT 11
 AAC85748
 ID AAC85748 standard; cDNA; 1158 BP.
 XX
 AC AAC85748;
 XX
 DT 18-JUL-2001 (first entry)
 XX
 DE Mutant delta 12 alpha-galactosidase A cDNA.
 XX
 DE Wild type; human; alpha-galactosidase A; alpha-Gal-A; variant; deletion;
 KW Fabry disease; C-terminal truncation; ss.
 KW Homo sapiens.
 OS
 XX US6210666-B1.
 XX
 XX 03-APR-2001.
 XX
 XX 21-OCT-1998; 98US-00176666.
 XX
 XX 21-OCT-1997; 97US-0062560P.
 PR
 PA (ORPH-) ORPHAN MEDICAL INC.
 XX
 XX Miyamura N;
 PI
 XX WPI; 2001-289627/30.
 DR
 DR P-PSDB; AAB47203.
 XX
 XX New alpha-galactosidase A polypeptide, useful for treating Fabry disease
 PT in humans or for treating a condition associated with a deficiency of
 PT alpha-galactosidase A.
 XX
 XX Disclosure; Col 71-74; 51pp; English.

XX This sequence encodes a variant human alpha-galactosidase A (alpha-Gal-A)
 CC which has a C-terminal truncation of 12 amino acids. Alpha-Gal A or its
 CC variant, which has a carboxy-terminal deletion of 2-11 amino acids and
 CC which exhibits alpha-galactosidase A enzyme activity, is useful for
 CC treating Fabry disease or a condition associated with a deficiency of
 CC alpha-galactosidase A. Fabry disease has been linked to an insertion of
 CC an adenine nucleotide at nt 1095 in exon 7 of the alpha-Gal A coding
 CC sequence, which results in a tyrosine-to-stop codon substitution at amino
 CC acid residue 365 (Y365X), causing a C-terminal truncation by 65 residues
 XX

SQ Sequence 1158 BP; 299 A; 259 C; 301 G; 299 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 6.85e-06 Length: 1158
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-10-602-220-29 (1-18) x AAC85748 (1-1158)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 |||||
 Db 1105 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCAGCTA 1158
 |||||

```
RESULT 12
AAC85747
ID AAC85747 standard; cDNA; 1161 BP.
XX
AC AAC85747;
XX
DT 18-JUL-2001 (first entry)
XX
DE Mutant delta 11 alpha-galactosidase A cDNA.
XX
KW Wild type; human; alpha-galactosidase A; alpha-Gal-A; variant; deletion;
KW Fabry disease; C-terminal truncation; ss.
XX
OS Homo sapiens.
XX
FN US6210666-B1.
XX
PD 03-APR-2001.
XX
PF 21-OCT-1998; 98US-00176666.
XX
PR 21-OCT-1997; 97US-0062560P.
XX
PA (ORPH-) ORPHAN MEDICAL INC.
XX
PI Miyamura N;
XX
DR WPI; 2001-289627/30.
XX
PD P-PSDB; AAB47201.
XX
PF New alpha-galactosidase A polypeptide, useful for treating Fabry disease
PF in humans or for treating a condition associated with a deficiency of
PR alpha-galactosidase A.
XX
PS Disclosure; Col 71-72; 51pp; English.
XX
PI Miyamura N;
XX
DR WPI; 2001-289627/30.
XX
PD P-PSDB; AAB47202.
XX
PF New alpha-galactosidase A polypeptide, useful for treating Fabry disease
PF in humans or for treating a condition associated with a deficiency of
PR alpha-galactosidase A.
XX
PS Disclosure; Col 71-72; 51pp; English.
XX
CC This sequence encodes a variant human alpha-galactosidase A (alpha-Gal-A)
CC which has a C-terminal truncation of 11 amino acids. Alpha-Gal A or its
CC variant, which has a carboxy-terminal deletion of 2-11 amino acids and
CC which exhibits alpha-galactosidase A enzyme activity, is useful for
CC treating Fabry disease or a condition associated with a deficiency of
CC alpha-galactosidase A. Fabry disease has been linked to an insertion of
CC an adenine nucleotide at nt 1095 in exon 7 of the alpha-Gal A coding
CC sequence, which results in a tyrosine-to-stop codon substitution at amino
CC acid residue 365 (Y365X), causing a C-terminal truncation by 65 residues
XX
SQ Sequence 1161 BP; 301 A; 259 C; 302 G; 299 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 6.87e-06 Length: 1161
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-602-220-29 (1-18) x AAC85747 (1-1161)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 1105 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGTTCAGCTA 1158

RESULT 13
AAC85746
ID AAC85746 standard; cDNA; 1164 BP.
XX
AC AAC85746;
XX
DT 18-JUL-2001 (first entry)
XX
DE Mutant delta 10 alpha-galactosidase A cDNA.
XX
KW Wild type; human; alpha-galactosidase A; alpha-Gal-A; variant; deletion;
KW Fabry disease; C-terminal truncation; ss.
XX
OS Homo sapiens.
XX
FN US6210666-B1.
XX
PD 03-APR-2001.
XX
PF 21-OCT-1998; 98US-00176666.
XX
PR 21-OCT-1997; 97US-0062560P.
XX
PA (ORPH-) ORPHAN MEDICAL INC.
XX
PI Miyamura N;
XX
DR WPI; 2001-289627/30.
XX
PD P-PSDB; AAB47201.
XX
PF New alpha-galactosidase A polypeptide, useful for treating Fabry disease
PF in humans or for treating a condition associated with a deficiency of
PR alpha-galactosidase A.
XX
PS Disclosure; Col 69-72; 51pp; English.
XX
CC This sequence encodes a variant human alpha-galactosidase A (alpha-Gal-A)
CC which has a C-terminal truncation of 10 amino acids. Alpha-Gal A or its
CC variant, which has a carboxy-terminal deletion of 2-11 amino acids and
CC which exhibits alpha-galactosidase A enzyme activity, is useful for
CC treating Fabry disease or a condition associated with a deficiency of
CC alpha-galactosidase A. Fabry disease has been linked to an insertion of
CC an adenine nucleotide at nt 1095 in exon 7 of the alpha-Gal A coding
CC sequence, which results in a tyrosine-to-stop codon substitution at amino
CC acid residue 365 (Y365X), causing a C-terminal truncation by 65 residues
XX
SQ Sequence 1164 BP; 303 A; 259 C; 302 G; 300 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 6.89e-06 Length: 1164
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-602-220-29 (1-18) x AAC85746 (1-1164)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 1105 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGTTCAGCTA 1158

RESULT 14
AAC85745
ID AAC85745 standard; cDNA; 1167 BP.
XX
AC AAC85745;
XX
DT 18-JUL-2001 (first entry)
XX
DE Mutant delta 9 alpha-galactosidase A cDNA.
XX
KW Wild type; human; alpha-galactosidase A; alpha-Gal-A; variant; deletion;
KW Fabry disease; C-terminal truncation; ss.
XX
OS Homo sapiens.
XX
FN US6210666-B1.
XX
PD 03-APR-2001.
XX
PF 21-OCT-1998; 98US-00176666.
XX
```

```
PR 21-OCT-1997; 97US-00625560P.
XX (ORPH-) ORPHAN MEDICAL INC.
XX Miyamura N;
XX WPI; 2001-289627/30.
XX P-PSDB; AAB47200.
XX New alpha-galactosidase A polypeptide, useful for treating Fabry disease
PT in humans or for treating a condition associated with a deficiency of
PT alpha-galactosidase A.
XX Disclosure; Col 67-70; 51pp; English.
XX This sequence encodes a variant human alpha-galactosidase A (alpha-Gal-A)
CC which has a C-terminal truncation of 9 amino acids. Alpha-Gal A or its
CC variant, which has a carboxy-terminal deletion of 2-11 amino acids and
CC which exhibits alpha-galactosidase A enzyme activity, is useful for
CC treating Fabry disease or a condition associated with a deficiency of
CC alpha-galactosidase A. Fabry disease has been linked to an insertion of
CC an adenine nucleotide at nt 1095 in exon 7 of the alpha-Gal A coding
CC sequence, which results in a tyrosine-to-stop codon substitution at amino
CC acid residue 365 (Y365X), causing a C-terminal truncation by 65 residues
XX SQ Sequence 1167 BP; 305 A; 260 C; 302 G; 300 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 6.91e-06 Length: 1167
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-602-220-29 (1-18) x AAC85745 (1-1167)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAGAGAGTCACATAAATCCCAAGGCACCTGTTTGGCTTCAGCTA 1158

RESULT 15
AAC85744
ID AAC85744 standard; cDNA; 1170 BP.
XX AC
XX AAC85744;
XX 18-JUL-2001 (first entry)
XX DE Mutant delta 8 alpha-galactosidase A cDNA.
XX Wild type; human; alpha-galactosidase A; alpha-Gal-A; variant; deletion;
XX KW Fabry disease; C-terminal truncation; ss.
XX OS Homo sapiens.
XX EN US6210666-B1.
XX PD 03-APR-2001.
XX PF 21-OCT-1998; 98US-00176666.
XX PR 21-OCT-1997; 97US-00625560P.
XX (ORPH-) ORPHAN MEDICAL INC.
XX PA Miyamura N;
XX PI
XX WPI; 2001-289627/30.
XX DR P-PSDB; AAB47199.
XX New alpha-galactosidase A polypeptide, useful for treating Fabry disease
PT in humans or for treating a condition associated with a deficiency of
```

```
PT alpha-galactosidase A.
XX Disclosure; Col 67-68; 51pp; English.
XX This sequence encodes a variant human alpha-galactosidase A (alpha-Gal-A)
CC which has a C-terminal truncation of 8 amino acids. Alpha-Gal A or its
CC variant, which has a carboxy-terminal deletion of 2-11 amino acids and
CC which exhibits alpha-galactosidase A enzyme activity, is useful for
CC treating Fabry disease or a condition associated with a deficiency of
CC alpha-galactosidase A. Fabry disease has been linked to an insertion of
CC an adenine nucleotide at nt 1095 in exon 7 of the alpha-Gal A coding
CC sequence, which results in a tyrosine-to-stop codon substitution at amino
CC acid residue 365 (Y365X), causing a C-terminal truncation by 65 residues
XX SQ Sequence 1170 BP; 306 A; 260 C; 303 G; 301 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 6.94e-06 Length: 1170
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-602-220-29 (1-18) x AAC85744 (1-1170)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAGAGAGTCACATAAATCCCAAGGCACCTGTTTGGCTTCAGCTA 1158

Search completed: January 1, 2006, 00:49:55
Job time : 51.2483 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2006, 00:27:17 ; Search time 326.524 Seconds
(without alignments)
2579.189 Million cell updates/sec

Title: US-10-602-220-29

Perfect score: 89
Sequence: 1 TSLRSHINPTGTVLLQL 18

Scoring table: BLOSUM62 Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO spoil/US10602220/runat_30122005_140558_5433/app_query.fasta_1.782
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0 -LOECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10602220@cgn 1.1.6731 @runat_30122005_140558_5433 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hic:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gsei:*
10: gb_gsei2:*
11: gb_gsei3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	89	100.0	177	1	AI6211155
C 2	89	100.0	311	6	CD365384
C 3	89	100.0	319	2	BP061865
C 4	89	100.0	346	1	AA251784
C 5	89	100.0	358	1	AW948701
C 6	89	100.0	371	3	BI492391
C 7	89	100.0	372	1	AW021436

C 8	89	100.0	387	1	AI758588
C 9	89	100.0	393	1	AW004664
C 10	89	100.0	397	1	AA287599
C 11	89	100.0	398	1	AW449396
C 12	89	100.0	401	1	AI948576
C 13	89	100.0	403	1	AI949177
C 14	89	100.0	407	1	AA548321
C 15	89	100.0	427	1	AIU149911
C 16	89	100.0	443	1	AW470828
C 17	89	100.0	445	1	AW135099
C 18	89	100.0	446	6	CD742642
C 19	89	100.0	460	6	CA777417
C 20	89	100.0	461	2	BF001466
C 21	89	100.0	462	2	BF111590
C 22	89	100.0	462	5	BU740166
C 23	89	100.0	481	1	AI799133
C 24	89	100.0	496	1	AI708415
C 25	89	100.0	499	1	AA708506
C 26	89	100.0	509	1	AA622854
C 27	89	100.0	512	1	AI859628
C 28	89	100.0	521	1	AA664321
C 29	89	100.0	524	2	BG232101
C 30	89	100.0	530	6	CB529984
C 31	89	100.0	542	1	AI469893
C 32	89	100.0	544	1	AW731746
C 33	89	100.0	544	2	BF590466
C 34	89	100.0	547	2	BE855457
C 35	89	100.0	547	2	BE855462
C 36	89	100.0	551	3	BP299850
C 37	89	100.0	554	1	AA463854
C 38	89	100.0	554	5	BU615633
C 39	89	100.0	568	3	BP299872
C 40	89	100.0	575	1	AI800468
C 41	89	100.0	577	1	AW517319
C 42	89	100.0	593	1	AW264095
C 43	89	100.0	600	6	CD367153
C 44	89	100.0	601	5	BU607709
C 45	89	100.0	602	1	AI859634

ALIGNMENTS

RESULT 1
AI621155/c
LOCUS
DEFINITION
tu88b03.x1 NCI CGAP Gas4 Homo sapiens cDNA clone IMAGE:2258093 3', similar to gb:X14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 177)
AUTHORS
TITLE
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL
COMMENT
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco.
Location/Qualifiers

```

source
1. 177
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2258093"
/tissue_type="poorly differentiated adenocarcinoma with
  signet ring cell features"
/lab_host="DH10B"
/clone_lib="NCI CGAP Gas4"
/note="Organ: stomach; Vector: pCMV-SPORT6; Site 1: SalI;
  Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
  Average insert size 1.69 kb. Life Technologies catalog #:
  11549-011"

ORIGIN
Alignment Scores:
Pred. No.: 2.8e-05 Length: 177
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-602-220-29 (1-18) x A1621155 (1-177)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 96 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTGTTCAGCTA 43

RESULT 2
LOCUS CD365384 311 bp mRNA linear EST 05-AUG-2004
DEFINITION UI-H-FT2-bj1-d-03-0-UI.s1 NCI CGAP_Ft2 Homo sapiens cDNA clone
similar to SW:AGAL_HUMAN P06280 ALPHA-GALACTOSIDASE A PRECURSOR ;,
mRNA sequence.
ACCESSION CD365384.1 GI:31149474
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 311)
NCI CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/cgap.html
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
source
1. 311
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-FT2-bj1-d-03-0-UI"
/tissue_type="Alveolar Macrophage"
/dev stage="Adult"
/lab host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP_Ft2"
/note="Organ: Lung; Vector: pTT3-Pac (Pharmacia) with a
  modified polylinker; Site 1: EcoR I; Site 2: Not I;
  NCI CGAP_Ft2 is a subcloned cDNA library constructed from
  a pool of 81 RNA samples from Alveolar Macrophages
  challenged with different treatments. The mRNA samples
  were a mixture of these conditions (times refer to
  incubations following isolation by bronchoalveolar lavage)
  (some normal donor macrophages were cultured in some of
  the conditions, other donor macrophages in different
  conditions). The mRNA samples were pooled for library
  construction. Control 0 hours; control 3 hours; control 24
  hours; LPS 100 ng/ml, 3 hours; LPS 100 ng/ml, 24 hours;
  PMA 10 ng/ml, 3 hours; PMA 10 ng/ml, 24 hours; Klebsiella
  moi 10, 3 hours; Klebsiella moi 10, 24 hours; Staph
  aureus moi 10, 3 hours; Staph aureus moi 10, 24 hours;
  Adenoviral vector (Ad5 CMV eGFP), moi 500, 3 hours; wt
  Adenoviral vector (Ad5 CMV eGFP), moi 500, 24 hours; wt
  Adenovirus moi 500, 3 hours; wt Adenovirus moi 500, 24
  hours; Ad vector + LPS 3 hours; Ad vector + LPS 24 hours;
  wt Adenovirus + LPS 3 hours; wt Adenovirus + LPS 24 hours.
  The library was subtracted according to Bonaldo, Lennon
  and Soares, Genome Research, 6:791-806, 1996. The tissue
  was provided by Dr. Gary W. Hunninghake of the University
  of Iowa.
  TAG_TISSUE=Human Lung Alveolar Macrophage
  TAG_LIB=UI-H-FT2
  TAG_SEQ=GGCCATGCCG"

ORIGIN
Alignment Scores:
Pred. No.: 5.31e-05 Length: 311
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x CD365384 (1-311)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 109 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTGTTCAGCTA 56

RESULT 3
LOCUS BF061865 319 bp mRNA linear EST 16-OCT-2000
DEFINITION 7k68a08.x1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:3480471.3
similar to SW:AGAL_HUMAN P06280 ALPHA-GALACTOSIDASE A PRECURSOR ;,
mRNA sequence.
ACCESSION BF061865
VERSION BF061865.1 GI:10820775
KEYWORDS EST.
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
1 (bases 1 to 319)
NCI CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco.
Location/Qualifiers
1. 319
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

FEATURES
source
1. 319
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

```

/clone="IMAGE:3480471"
 /tissue type="pooled germ cell tumors"
 /lab host="DH10B"

/clone lib="NCI CGAP GC6"
 /note="Vector: pT7T3P-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIda 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Alignment Scores:
 Pred. No.: 5,46e-05 Length: 319
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-10-602-220-29 (1-18) x BF061865 (1-319)

Oy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 |||||
 Db 102 ACTTCAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTGGTTCAGCTA 49

RESULT 4

AA251784/c
 LOCUS AA251784 346 bp mRNA linear EST 13-AUG-1997
 DEFINITION zsl1d08.s1 NCI CGAP GCBI Homo sapiens cDNA clone IMAGE:684879 3' similar to gb:X14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN); mRNA sequence.

ACCESSION

AA251784

VERSION

AA251784.1 GI:1886746

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 346)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LNL ; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Seq primer: 41ml3 fwd. ET from Amersham

High quality sequence stop: 307.

FEATURES

Location/Qualifiers

1..346
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:684879"
 /tissue_type="germinal center B cell"
 /lab host="DH10B"
 /clone lib="NCI CGAP GCBI"

/note="Vector: pT7T3P-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer

[5'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTTTT-3'. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Alignment Scores:
 Pred. No.: 5,99e-05 Length: 346
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 1 Gaps: 0

US-10-602-220-29 (1-18) x AA251784 (1-346)

Oy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 |||||
 Db 100 ACTTCAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTGGTTCAGCTA 47

RESULT 5

AW948701

LOCUS

AW948701

DEFINITION

OVO-FT0002-050500-231-h02 FT0002 Homo sapiens cDNA, mRNA sequence.

ACCESSION

AW948701

VERSION

AW948701.1 GI:8126475

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 358)

REFERENCE

1 (bases 1 to 358)

AUTHORS

Nagai, M.A., da Silva, W. Jr., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

PUBMED

10737800

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=st2=QV0-FT0002-050>)

500-231-h02st3=2000-05-05st4=1)

Seq primer: puc 18 forward

High quality sequence start: 10

High quality sequence stop: 358.

FEATURES

Location/Qualifiers

1..358
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev stage="Adult"

/clone lib="FT0002"

/note="Organ: prostate tumor; Vector: puc18; Site 1: SmaI;

Site 2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (O.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

ORIGIN

```

Alignment Scores:
Pred. No.: 6.23e-05 Length: 358
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-602-220-29 (1-18) x AW948701 (1-358)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 279 ACTTCAGGTTAAGAGTCACATATATCCACAGGCACTGTTTTCAGCTA 332

RESULT 6
LOCUS BI492391/c 371 bp mRNA linear EST 28-AUG-2001
DEFINITION IMAGE:2484240 3', mRNA sequence.
ACCESSION BI492391
VERSION BI492391.1 GI:15331735
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 371)
AUTHORS Robertson,N.G., Khetarpal,U., Gutierrez-Espeleta,G.A., Bieber,F.R.
and Morton,C.C.
TITLE Isolation of novel and known genes from a human fetal cochlear cDNA
library using subtractive hybridization and differential screening
JOURNAL Genomics 23, 42-50 (1994)
PUBMED 7829101
COMMENT Contact: Morton, C. C.
Departments of Pathology and Obstetrics, Gynecology and
Reproductive Biology
Brigham and Women's Hospital
75 Francis Street, Harvard Medical School, Boston, MA 02115, USA
Tel: 617 732 7980
Fax: 617 738 6996
Email: ccmorton@rics.bwh.harvard.edu
DNA sequencing and analyses were performed by National Institutes
of Health Intramural Sequencing Center (NISC; see
http://www.nisc.nih.gov).
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Plate: LLAM6181 row: L column: 1
Seq primer: P7 primer.
FEATURES
source
location/Qualifiers
1..371
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2484240"
/tissue_type="cochlea"
/dev_stage="16-22 week fetus"
/lab_host="SOLR cells (kanamycin resistant)"
/clone_lib="Morton Fetal Cochlea"
/note="Organ: ear; Vector: pBluescript SK-; Site 1: ECoRI;
Site 2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned
unidirectionally. Primer: Oligo dt. Fetal cochlea, normal.
37% of inserts <0.5 Kb, 56% 0.5-1.0 Kb, 7% >1 Kb. Uni-ZAP
XR Vector. Library constructed by N. Robertson, C. Morton.
-5' adaptor sequence: 5' GAATTCGGCAGGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

ORIGIN
Alignment Scores:
Pred. No.: 6.49e-05 Length: 371
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 115 ACTTCAGGTTAAGAGTCACATATATCCACAGGCACTGTTTTCAGCTA 62

RESULT 7
LOCUS AW021436 372 bp mRNA linear EST 13-SEP-1999
DEFINITION IMAGE:2484240 5', mRNA sequence.
ACCESSION AW021436
VERSION AW021436.1 GI:5874966
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 372)
AUTHORS Robertson,N.G., Khetarpal,U., Gutierrez-Espeleta,G.A., Bieber,F.R.
and Morton,C.C.
TITLE Isolation of novel and known genes from a human fetal cochlear cDNA
library using subtractive hybridization and differential screening
JOURNAL Genomics 23, 42-50 (1994)
PUBMED 7829101
COMMENT Contact: Morton, C. C.
Departments of Pathology and Obstetrics, Gynecology and
Reproductive Biology
Brigham and Women's Hospital
75 Francis Street, Harvard Medical School, Boston, MA 02115, USA
Tel: 617 732 7980
Fax: 617 738 6996
Email: ccmorton@rics.bwh.harvard.edu
DNA sequencing and analyses were performed by National Institutes
of Health Intramural Sequencing Center (NISC; see
http://www.nisc.nih.gov).
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Plate: LLAM6181 row: L column: 1
Seq primer: M13RP1 reverse primer (ABI).
FEATURES
source
location/Qualifiers
1..372
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2484240"
/tissue_type="cochlea"
/dev_stage="16-22 week fetus"
/lab_host="SOLR cells (kanamycin resistant)"
/clone_lib="Morton Fetal Cochlea"
/note="Organ: ear; Vector: pBluescript SK-; Site 1: ECoRI;
Site 2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned
unidirectionally. Primer: Oligo dt. Fetal cochlea, normal.
37% of inserts <0.5 Kb, 56% 0.5-1.0 Kb, 7% >1 Kb. Uni-ZAP
XR Vector. Library constructed by N. Robertson, C. Morton.
-5' adaptor sequence: 5' GAATTCGGCAGGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"

ORIGIN
Alignment Scores:
Pred. No.: 6.51e-05 Length: 372
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-602-220-29 (1-18) x AW021436 (1-372)

```

```

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 257 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGTTCAGCTA 310

RESULT 8
LOCUS AW004664/c
DEFINITION 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
similar to gb:U14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN) ;
mRNA sequence.

ACCESSION AI758588
VERSION AI758588.1 GI:5152313
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 387)
REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1375 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 217.
Location/Qualifiers
1..387
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2278435"
/tissue_type="poorly-differentiated endometrial
adenocarcinoma, 2 pooled tumors"
/lab_host="DH10B"
/clone_lib="NCI CGAP Ut3"
/notes="Organ: uterus; Vector: pCMV-SPORT6; Site 1: Salt;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.45 kb. Life Technologies catalog #:
11541-018"

ORIGIN
Alignment Scores:
Pred. No.: 6,81e-05 Length: 387
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-602-220-29 (1-18) x AI758588 (1-387)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 91 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGTTCAGCTA 38

RESULT 9
LOCUS AW004664/c
DEFINITION 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
similar to gb:U14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN) ;
mRNA sequence.

ACCESSION AW004664
VERSION AW004664.1 GI:5853532
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 393)
REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,
M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 556 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 332.
Location/Qualifiers
1..393
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2505343"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/clone_lib="NCI CGAP Co3"
/notes="Vector: pTT3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from 12 pooled bulk tumor samples and primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pTT3 vector. Library went through one round of
normalization."

ORIGIN
Alignment Scores:
Pred. No.: 6,93e-05 Length: 393
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-602-220-29 (1-18) x AW004664 (1-393)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 93 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGTTCAGCTA 40

RESULT 10
LOCUS AA287599/c
DEFINITION AA287599 397 bp mRNA linear EST 15-AUG-1997
similar to gb:U14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN) ;
mRNA sequence.

ACCESSION AA287599
VERSION AA287599.1 GI:1933299
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

```

```
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 397)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 2004 Std Error: 0.00
Seq primer: -41ml3 fwd. Et from Amersham
High quality sequence stop: 381.
Location/Qualifiers
1. 397
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:5740423"
/db_xref="taxon:9606"
/clone="IMAGE:700989"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/clone_lib="NCI CGAP GCBI"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTCGAGCGCGCTCATTTTCTTTT-3'
1. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN
Alignment Scores:
Pred. No.: 7,01e-05 Length: 397
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-602-220-29 (1-18) x AA287599 (1-397)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
|||||
Db 100 ACTTCACGGTTAAGAAGTCACATAATCCACAGGCACCTGTTTCTTCAGCTA 47

RESULT 11
AW449396/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AW449396
UI-H-B13-aki-b-05-0-UI.e1 NCI CGAP_Sub5 Homo sapiens cDNA clone
IMAGE:2734688 3', mRNA sequence.
AW449396
AW449396.1 GI:6990172
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 398)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
```

strand cDNA and therefore this may represent a bonafide poly A tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www-bio.lnl.gov/bbrp/image/image.html Seq primer: M13 Forward POLYA=Yes.

FEATURES

source

Location/Qualifiers

```
1. 398
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2734688"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP Sub5"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; NCI CGAP Sub5
is a subtracted library derived from NCI CGAP_Sub4. The
NCI CGAP Sub5 library had 3 million recombinants. A
single-stranded DNA preparation of NCI CGAP Sub4 was used
as a tracer in a subtractive hybridization with a driver
comprising the IMAGE pool (NCI CGAP Kid3 pool 1 LLM
3334-3337, 3682-3683, 3798-3803 (IMAGE CloneIDs
1323376-1323911, 1456008-1456775, 1500552-1502855);
NCI CGAP Kid5 pool 1 LLM 3338-3342, 3722-3725, 3776-3778
(IMAGE CloneIDs 1323912-1325831, 1471368-1472903,
1492104-1493255); NCI CGAP_Lu5 pool 1 LLM 3575-3582,
3851-3854 (IMAGE CloneIDs
1414920-1417991, 1520904-1522439); NCI CGAP_GC4 pool 1 LLM
3164-3167, 3716-3720, 3733-3735 (IMAGE CloneIDs
1257096-1258631, 1469064-1470983, 1475592-1476743);
NCI CGAP Pr22 pool 1 LLM 2457-2459, 2758-2759, 3062-3068
(IMAGE CloneIDs 985608-986759, 1101192-1101959,
1217928-1220615); NCI CGAP_Col0 pool 1 LLM 2644-2653,
2871-2872 (IMAGE CloneIDs
1057416-1061255, 1144584-1145351). (10% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI CGAP Sub1 (IMAGE CloneIDs 2708616-2710535) and
NCI CGAP Sub2 (IMAGE CloneIDs 2710536-2712455) (10% of
the driver population), plus a pool of 11,136 clones from
NCI CGAP Sub3 (IMAGE CloneIDs 2712456-2723591) (10% of the
driver population), plus a pool of 5,472 clones from
NCI CGAP Sub4 (IMAGE CloneIDs 2723592-2728969) (70% of the
driver population). Subtraction was performed as
previously described [Bonaldo, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6, 791-806.
TAG_TISSUE=kidney
TAG_LIIB=NCI CGAP_Kids
TAG_SEQ=ATTC"
```

ORIGIN

Alignment Scores:

```
Pred. No.: 7,03e-05 Length: 398
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
```

US-10-602-220-29 (1-18) x AW449396 (1-398)

```
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
|||||
Db 108 ACTTCACGGTTAAGAAGTCACATAATCCACAGGCACCTGTTTCTTCAGCTA 55
```

RESULT 12

LOCUS

AI948576/c

DEFINITION

similar to gb:U14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN);

mRNA sequence.

ACCESSION

AI948576

AI948576 401 bp mRNA linear EST 08-MAR-2000
wq07b03.x1 NCI CGAP Kid12 Homo sapiens cDNA clone IMAGE:2470541 3',
similar to gb:U14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN);,
mRNA sequence.
AI948576

```

VERSION
KEYWORDS
SOURCE
ORGANISM
A1948576.1 GI:5740886
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 401)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 466 Std Error: 0.00
Seq primer: -40UP from Gibco.
Location/Qualifiers
1..401
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2470541"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/clone_lib="NCI CGAP Kid12"
/note="Organ: Kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not 1; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP Kid5 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneDs 1323912-1325831, 1471368-1472903 and
1492104-1493255). Subtraction by Bento Soares and M.
Fatima Bonaldo."
FEATURES
source
Location/Qualifiers
1..401
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2470541"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/clone_lib="NCI CGAP Kid12"
/note="Organ: Kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not 1; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP Kid5 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneDs 1323912-1325831, 1471368-1472903 and
1492104-1493255). Subtraction by Bento Soares and M.
Fatima Bonaldo."
ORIGIN
Alignment Scores:
Pred. No.: 7,096-05 Length: 401
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
US-10-602-220-29 (1-18) x A1948576 (1-401)
Oy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
Db 94 ACTTCAAGGTTAAGAGTCACATTAATCCACAGGCACCTGTTTGCCTCAGCTA 41
RESULT 13
A1949177/c
LOCUS
DEFINITION
wq08h03.x1 NCI CGAP Kid12 Homo sapiens cDNA clone IMAGE:2470709 3',
similar to gb:X14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN);,
mRNA sequence.
ACCESSION
A1949177
VERSION
A1949177.1 GI:5741487
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 427)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 496 Std Error: 0.00
Seq primer: -40UP from Gibco.
Location/Qualifiers
1..403
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2470709"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/clone_lib="NCI CGAP Kid12"
/note="Organ: Kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not 1; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP Kid5 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneDs 1323912-1325831, 1471368-1472903 and
1492104-1493255). Subtraction by Bento Soares and M.
Fatima Bonaldo."
ORIGIN
Alignment Scores:
Pred. No.: 7,136-05 Length: 403
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
US-10-602-220-29 (1-18) x A1949177 (1-403)
Oy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
Db 96 ACTTCAAGGTTAAGAGTCACATTAATCCACAGGCACCTGTTTGCCTCAGCTA 43
RESULT 14
AAS48321/c
LOCUS
DEFINITION
nk16e01.81 NCI CGAP Coll Homo sapiens cDNA clone IMAGE:1013688 3',
similar to gb:X14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN);,
mRNA sequence.
ACCESSION
AAS48321
VERSION
AAS48321.1 GI:2318603
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 427)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
```

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Elias Campo, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Stratagene, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.lnl.gov/bbrp/image.html
Insert Length: 1118 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 364.

FEATURES
source
1. .427
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1013688"
/tissue type="tumor"
/lab host="SOLR (kanamycin resistant)"
/clone lib="NCI CGAP Coll"
/note="Organ: colon; Vector: Bluescript SK-; Site:1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Multiple colon tumors. 5' adaptor sequence: 5' GAATTCGGCAGCAG 3' 3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

ORIGIN

Alignment Scores:
Pred. No.: 7.61e-05 Length: 427
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-602-220-29 (1-18) x AA548321 (1-427)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 90 ACTTCAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTA 37

FEATURES
source
1. .437
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP2000908"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/clone lib="NT2RP2"
/note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor cells after 2-weeks retinoic acid (RA) induction"

ORIGIN

Alignment Scores:
Pred. No.: 7.82e-05 Length: 437
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-602-220-29 (1-18) x AU149911 (1-437)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 91 ACTTCAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTA 38

Search completed: January 1, 2006, 05:47:06
Job time : 330.524 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2006, 00:29:37 ; Search time 14.6483 Seconds
(without alignments)
2184.292 Million cell updates/sec

Title: US-10-602-220-29
Perfect score: 89
Sequence: 1 TSRLRSHNPRTGTVLLQL 18

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10602220/runat 30122005 140559 5451/app_query.fasta_1.782
-DB=Issued_Patents_NA -QPMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blowsum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10602220@cgn 1 1 237 @runat 30122005 140559 5451 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/1 COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5 COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/H COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PPCTUS COMB.seq.*
7: /cgn2_6/ptodata/1/ina/PP COMB.seq.*
8: /cgn2_6/ptodata/1/ina/RE COMB.seq.*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	100.0	210	3	US-08-928-881-19
2	89	100.0	210	3	US-09-543-921-19
3	89	100.0	210	3	US-09-266-014-1
4	89	100.0	210	3	US-09-491-759-19
5	89	100.0	297	2	US-07-602-824A-8
6	89	100.0	297	2	US-07-983-451-8
7	89	100.0	297	2	US-08-261-577-11
8	89	100.0	1158	3	US-09-176-666-53
9	89	100.0	1161	3	US-09-176-666-52

10	89	100.0	1164	3	US-09-176-666-51
11	89	100.0	1167	3	US-09-176-666-50
12	89	100.0	1170	3	US-09-176-666-49
13	89	100.0	1173	3	US-09-176-666-48
14	89	100.0	1176	3	US-09-176-666-47
15	89	100.0	1179	3	US-09-176-666-46
16	89	100.0	1182	3	US-09-176-666-45
17	89	100.0	1188	3	US-09-176-666-44
18	89	100.0	1194	3	US-09-176-666-43
19	89	100.0	1197	3	US-08-928-881-25
20	89	100.0	1197	3	US-09-543-921-25
21	89	100.0	1197	3	US-09-266-014-25
22	89	100.0	1197	3	US-09-491-759-25
23	89	100.0	1214	3	US-09-023-655-1001
24	89	100.0	1254	3	US-09-993-059-15
25	89	100.0	1254	3	US-10-103-327-15
26	89	100.0	1266	3	US-09-993-059-11
27	89	100.0	1266	3	US-10-103-327-11
28	89	100.0	1272	3	US-09-993-059-17
29	89	100.0	1278	3	US-10-103-327-17
30	89	100.0	1278	3	US-09-993-059-7
31	89	100.0	1278	3	US-10-103-327-7
32	89	100.0	1284	3	US-09-993-059-13
33	89	100.0	1284	3	US-10-103-327-13
34	89	100.0	1290	3	US-09-993-059-3
35	89	100.0	1296	3	US-10-103-327-3
36	89	100.0	1296	3	US-09-993-059-9
37	89	100.0	1296	3	US-10-103-327-9
38	89	100.0	1304	9	5179023-3
39	89	100.0	1308	3	US-09-993-059-5
40	89	100.0	1308	3	US-10-103-327-5
41	89	100.0	1343	3	US-08-928-881-18
42	89	100.0	1343	3	US-09-543-921-18
43	89	100.0	1343	3	US-09-266-014-3
44	89	100.0	1343	3	US-09-491-759-18
45	89	100.0	1393	2	US-07-602-824A-1

ALIGNMENTS

RESULT 1
US-08-928-881-19
; Sequence 19, Application US/08928881
; Patent No. 6083725
; GENERAL INFORMATION:
; APPLICANT: Selden et al., Richard F.
; TITLE OF INVENTION: THERAPY FOR ALPHA-GALACTOSIDASE A
; TITLE OF INVENTION: DEFICIENCY
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,881
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 07236/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:
LENGTH: 210 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-928-881-19

Alignment Scores:
Pred. No.: 2.06e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-08-928-881-19 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 155 ACTTCAAGGTTAAGAGTCACATAAATCCACAGGCACCTGTTTGTCTCAGCTA 208

RESULT 2

US-09-543-921-19
Sequence 19, Application US/09543921
Patent No. 6395884
GENERAL INFORMATION:
APPLICANT: Selden et al., Richard F.
TITLE OF INVENTION: THERAPY FOR ALPHA-GALACTOSIDASE A DEFICIENCY

NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/543,921
FILING DATE: 06-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/928,881
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 07236/003001
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906

INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 210 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 19:

US-09-543-921-19

Alignment Scores:
Pred. No.: 2.06e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-543-921-19 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 155 ACTTCAAGGTTAAGAGTCACATAAATCCACAGGCACCTGTTTGTCTCAGCTA 208

RESULT 3

US-09-266-014-1
Sequence 1, Application US/09266014
Patent No. 6458574

GENERAL INFORMATION:
APPLICANT: Selden, Richard F.
APPLICANT: Borowski, Marianne
APPLICANT: Kinoshita, Carol M
APPLICANT: Treco, Douglas A
APPLICANT: Williams, Melanie D
APPLICANT: Schuetz, Thomas J
APPLICANT: Daniel, Peter F.
TITLE OF INVENTION: Treatment for alpha-Galactosidase A Deficiency
FILE REFERENCE: FABRY DISEASE (18082-001)
CURRENT APPLICATION NUMBER: US/09/266,014
CURRENT FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: 60/026,041
PRIOR FILING DATE: 1996-09-13
PRIOR APPLICATION NUMBER: 08/928,881
PRIOR FILING DATE: 1997-09-12
PRIOR APPLICATION NUMBER: PCT US97/16603
PRIOR FILING DATE: 1997-09-12
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1
LENGTH: 210
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Human
OTHER INFORMATION: fibroblast library probe: exon 7, including
OTHER INFORMATION: amplification primers.

US-09-266-014-1

Alignment Scores:
Pred. No.: 2.06e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-266-014-1 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 155 ACTTCAAGGTTAAGAGTCACATAAATCCACAGGCACCTGTTTGTCTCAGCTA 208

RESULT 4

US-09-491-759-19
Sequence 19, Application US/09491759
Patent No. 6566099

GENERAL INFORMATION:
APPLICANT: Selden et al., Richard F.
TITLE OF INVENTION: THERAPY FOR ALPHA-GALACTOSIDASE A DEFICIENCY

NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/491,759
FILING DATE: 27-Jan-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/928,881
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 07236/003001
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 210 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 19:

US-09-491-759-19
Alignment Scores:
Pred. No.: 2,06e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-491-759-19 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
Db 155 ACTTCAAGGTTTAAAGAGTCACATAATCCACAGGCACTGTTTTCCTTCAGCTA 208

RESULT 5
US-07-602-824A-8
Sequence 8, Application US/07602824A
Patent No. 5356804
GENERAL INFORMATION:
APPLICANT: Desnick, Robert J.
APPLICANT: Bishop, David P.
APPLICANT: Ioannou, Yiannis A.
TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY
TITLE OF INVENTION: ACTIVE alpha-GALACTOSIDASE A
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/602,824A
FILING DATE: 24-OCT-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 6923-005
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..279
US-07-602-824A-8

Alignment Scores:
Pred. No.: 3.16e-07 Length: 297
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-10-602-220-29 (1-18) x US-07-602-824A-8 (1-297)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
Db 7 ACTTCAAGGTTTAAAGAGTCACATAATCCACAGGCACTGTTTTCCTTCAGCTA 60

RESULT 6
US-07-983-451-8
Sequence 8, Application US/07983451
Patent No. 5401650
GENERAL INFORMATION:
APPLICANT: Desnick, Robert J.
APPLICANT: Bishop, David P.
APPLICANT: Ioannou, Yiannis A.
TITLE OF INVENTION: Cloning and Expression of Biologically
TITLE OF INVENTION: Active alpha-Galactosidase A
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/983,451
FILING DATE: 30-NOV-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 07/983,451
REFERENCE/DOCKET NUMBER: 6923-030
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..279
US-07-983-451-8

Alignment Scores:
Pred. No.: 3.16e-07 Length: 297
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-10-602-220-29 (1-18) x US-07-983-451-8 (1-297)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
Db 7 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGAACTGTTTGGCTTCAGCTA 60

RESULT 7

US-08-261-577-11
; Sequence 11, Application US/08261577
; Patent No. 5580757
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; TITLE OF INVENTION: Cloning and Expression of Biologically
; TITLE OF INVENTION: Active alpha-Galactosidase A
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,577
; FILING DATE: 17-JUN-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-042
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 297 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 1..279

US-08-261-577-11

Alignment Scores:
Pred. No.: 3.16e-07 Length: 297
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-10-602-220-29 (1-18) x US-08-261-577-11 (1-297)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18

Db 7 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGAACTGTTTGGCTTCAGCTA 60

RESULT 8

US-09-176-666-53
; Sequence 53, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; TITLE OF INVENTION: FABRY DISEASE
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 53
; LENGTH: 1158
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-176-666-53

Alignment Scores:
Pred. No.: 1.69e-06 Length: 1158
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-176-666-53 (1-1158)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGAACTGTTTGGCTTCAGCTA 1158

RESULT 9

US-09-176-666-52
; Sequence 52, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; TITLE OF INVENTION: FABRY DISEASE
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 52
; LENGTH: 1161
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-176-666-52

Alignment Scores:
Pred. No.: 1.69e-06 Length: 1161
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-176-666-52 (1-1161)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGAACTGTTTGGCTTCAGCTA 1158

```
RESULT 10
US-09-176-666-51
; Sequence 51, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; EARLIER FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 51
; LENGTH: 1164
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-176-666-51

Alignment Scores:
Pred. No.: 1.7e-06 Length: 1164
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-176-666-51 (1-1164)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTCAGCTA 1158

RESULT 11
US-09-176-666-50
; Sequence 50, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; EARLIER FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 50
; LENGTH: 1167
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-176-666-50

Alignment Scores:
Pred. No.: 1.7e-06 Length: 1167
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-176-666-50 (1-1167)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTCAGCTA 1158

RESULT 12
US-09-176-666-49
; Sequence 49, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; EARLIER FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 49
; LENGTH: 1170
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-176-666-49

Alignment Scores:
Pred. No.: 1.71e-06 Length: 1170
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-176-666-49 (1-1170)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTCAGCTA 1158

RESULT 13
US-09-176-666-48
; Sequence 48, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; EARLIER FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 48
; LENGTH: 1173
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-176-666-48

Alignment Scores:
Pred. No.: 1.71e-06 Length: 1173
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-176-666-48 (1-1173)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTCAGCTA 1158

RESULT 14
US-09-176-666-47
; Sequence 47, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; EARLIER FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 47
; LENGTH: 1173
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-176-666-47

Alignment Scores:
Pred. No.: 1.71e-06 Length: 1173
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-176-666-47 (1-1173)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTCAGCTA 1158
```

Tue Jan 3 11:36:10 2006

; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; TITLE OF INVENTION: FERRY DISEASE

; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 47
; LENGTH: 1176
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-176-666-47

Alignment Scores:
Pred. No.: 1.72e-06 Length: 1176
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-176-666-47 (1-1176)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTCCTTCAGCTA 1158

RESULT 15

US-09-176-666-46
; Sequence 46, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; TITLE OF INVENTION: FERRY DISEASE
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 46
; LENGTH: 1179
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-176-666-46

Alignment Scores:
Pred. No.: 1.73e-06 Length: 1179
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-176-666-46 (1-1179)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTCCTTCAGCTA 1158

Search completed: January 1, 2006, 05:53:06
Job time : 15.6483 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2006, 03:35:43 ; Search time 64.4276 Seconds
(without alignments)
2310.327 Million cell updates/sec

Title: US-10-602-220-29
Perfect score: 89
Sequence: 1 TSRLRSHNPTGTVLLQL 18

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US10602220/runat_30122005_140601_5538/app_query.fasta_1.782
-DB=PublishedApplications_NA_Main -QFWT=fastap -SUFFIX=p2n.rnpbm
-MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=-1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10602220 CGEN 1.1.1364 @runat_30122005_140601_5538 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-FILE_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA_Main:

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89	100.0	210	5	US-10-165-060-1
2	89	100.0	210	5	US-10-165-968-1
3	89	100.0	210	6	US-10-318-905-19
4	89	100.0	227	7	US-10-242-535A-29035
5	89	100.0	227	7	US-10-085-783A-29035
6	89	100.0	270	7	US-10-242-535A-30531
7	89	100.0	270	7	US-10-085-783A-30531

ALIGNMENTS

RESULT 1
US-10-165-060-1
; Sequence 1, Application US/10165060
; Publication No. US20030077806A1
; GENERAL INFORMATION:
; APPLICANT: Selden, Richard F
; APPLICANT: Borowski, Marianne
; APPLICANT: Kinoshita, Carol M
; APPLICANT: Treco, Douglas A
; APPLICANT: Williams, Melanie D
; APPLICANT: Schuetz, Thomas J
; APPLICANT: Daniel, Peter F.
; TITLE OF INVENTION: Treatment for alpha-Galactosidase A Deficiency
; FILE REFERENCE: FABRY DISEASE (18082-001)
; CURRENT APPLICATION NUMBER: US/10/165,060
; CURRENT FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: US/09/266,014
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: 60/026,041
; PRIOR FILING DATE: 1996-09-13
; PRIOR APPLICATION NUMBER: 08/928,881
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: PCT US97/16603
; PRIOR FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 210
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Human
; OTHER INFORMATION: fibroblast library probe: exon 7, including
; OTHER INFORMATION: amplification primers.
US-10-165-060-1

Alignment Scores:
Pred. No.: 2.49e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-10-602-220-29 (1-18) x US-10-165-060-1 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 155 ACTTCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTGCTTCAGCTA 208

RESULT 2

US-10-165-968-1
; Sequence 1, Application US/10165968
; Publication No. US20030113894A1
; GENERAL INFORMATION:
; APPLICANT: Selden, Richard F
; APPLICANT: Borowski, Marianne
; APPLICANT: Kinoshita, Carol M
; APPLICANT: Treco, Douglas A
; APPLICANT: Williams, Melanie D
; APPLICANT: Schuetz, Thomas J
; APPLICANT: Daniel, Peter F.
; TITLE OF INVENTION: Treatment for alpha-Galactosidase A Deficiency
; FILE REFERENCE: FABRY DISEASE (18082-001)
; CURRENT APPLICATION NUMBER: US/10/165,968
; CURRENT FILING DATE: 2002-06-10
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: 60/026,041
; PRIOR FILING DATE: 1996-09-13
; PRIOR APPLICATION NUMBER: 08/928,881
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: PCT US97/16603
; PRIOR FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in ver. 2.0
; SEQ ID NO 1
; LENGTH: 210
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Human
; OTHER INFORMATION: fibroblast library probe: exon 7, including
; OTHER INFORMATION: amplification primers.
US-10-165-968-1

Alignment Scores:
Pred. No.: 2.49e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-10-602-220-29 (1-18) x US-10-165-968-1 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 155 ACTTCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTGCTTCAGCTA 208

RESULT 3

US-10-318-905-19
; Sequence 19, Application US/10318905
; Publication No. US20030152560A1

; GENERAL INFORMATION:
; APPLICANT: Selden et al., Richard F.
; TITLE OF INVENTION: THERAPY FOR ALPHA-GALACTOSIDASE A
; DEFICIENCY

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLING APPLICATION NUMBER: US/10/318,905

FILING DATE: 12-Dec-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/928,881

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Fraser, Janis K.

REGISTRATION NUMBER: 34,819

REFERENCE/DOCKET NUMBER: 07236/003001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/542-5070

TELEFAX: 617/542-8906

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 210 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 19:

US-10-318-905-19

Alignment Scores:

Pred. No.: 2.49e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x US-10-318-905-19 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 155 ACTTCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTGCTTCAGCTA 208

RESULT 4

US-10-242-535A-29035
; Sequence 29035, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28

; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29035
; LENGTH: 227
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (20)..(20)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-29035

Alignment Scores:
Pred. No.: 2.72e-07 Length: 227
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-10-602-220-29 (1-18) x US-10-242-535A-29035 (1-227)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 144 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGTTCAGCTA 197

RESULT 5

US-10-085-783A-29035
; Sequence 29035, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis

; FILE REFERENCE: 4231/2002
; CURRENT FILING DATE: 2002-02-28
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29035
; LENGTH: 227
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (20)..(20)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-29035

Alignment Scores:
Pred. No.: 2.72e-07 Length: 227
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-10-602-220-29 (1-18) x US-10-085-783A-29035 (1-227)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 144 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGTTCAGCTA 197

RESULT 6

US-10-242-535A-30531
; Sequence 30531, Application US/10242535A
; Publication No. US20040013663A1

; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30531
; LENGTH: 270
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (31)..(31)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-30531

Alignment Scores:
Pred. No.: 3.34e-07 Length: 270
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-10-602-220-29 (1-18) x US-10-242-535A-30531 (1-270)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 174 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGTTCAGCTA 227

RESULT 7

US-10-085-783A-30531
; Sequence 30531, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis

; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30531
; LENGTH: 270
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (31)..(31)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-30531

Alignment Scores:
Pred. No.: 3.34e-07 Length: 270
Score: 89.00 Matches: 18

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-10-602-220-29 (1-18) x US-10-085-783A-30531 (1-270)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 174 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTGGCTTCAGCTA 227

RESULT 8

US-10-165-060-5
; Sequence 5, Application US/10165060
; Publication No. US20030077806A1

GENERAL INFORMATION:

; APPLICANT: Selden, Richard F
; APPLICANT: Borowski, Marianne
; APPLICANT: Kinoshita, Carol M
; APPLICANT: Treco, Douglas A
; APPLICANT: Williams, Melanie D
; APPLICANT: Schuetz, Thomas J
; APPLICANT: Daniel, Peter F.

; TITLE OF INVENTION: Treatment for alpha-Galactosidase A Deficiency

; FILE REFERENCE: FABRY DISEASE (18082-001)

; CURRENT APPLICATION NUMBER: US/10/165,060

; CURRENT FILING DATE: 2002-06-07

; PRIOR APPLICATION NUMBER: US/09/266,014

; PRIOR FILING DATE: 1999-03-11

; PRIOR APPLICATION NUMBER: 60/026,041

; PRIOR FILING DATE: 1996-09-13

; PRIOR APPLICATION NUMBER: 08/928,881

; PRIOR FILING DATE: 1997-09-12

; PRIOR APPLICATION NUMBER: PCT US97/16603

; PRIOR FILING DATE: 1997-09-12

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 5

; LENGTH: 1197

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-165-060-5

Alignment Scores:
Pred. No.: 1.91e-06 Length: 1197
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-10-602-220-29 (1-18) x US-10-165-060-5 (1-1197)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTGGCTTCAGCTA 1158

RESULT 9

US-10-165-968-5

; Sequence 5, Application US/10165968

; Publication No. US20030113894A1

GENERAL INFORMATION:

; APPLICANT: Selden, Richard F
; APPLICANT: Borowski, Marianne
; APPLICANT: Kinoshita, Carol M
; APPLICANT: Treco, Douglas A
; APPLICANT: Williams, Melanie D
; APPLICANT: Schuetz, Thomas J
; APPLICANT: Daniel, Peter F.

; TITLE OF INVENTION: Treatment for alpha-Galactosidase A Deficiency

; FILE REFERENCE: FABRY DISEASE (18082-001)

; CURRENT APPLICATION NUMBER: US/10/165,968

; CURRENT FILING DATE: 2002-06-10

; PRIOR APPLICATION NUMBER: US/09/266,014
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: 60/026,041
; PRIOR FILING DATE: 1996-09-13
; PRIOR APPLICATION NUMBER: 08/928,881
; PRIOR FILING DATE: 1997-09-12

; PRIOR APPLICATION NUMBER: PCT US97/16603

; PRIOR FILING DATE: 1997-09-12

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 5

; LENGTH: 1197

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-165-968-5

Alignment Scores:
Pred. No.: 1.91e-06 Length: 1197
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-10-602-220-29 (1-18) x US-10-165-968-5 (1-1197)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTGGCTTCAGCTA 1158

RESULT 10

US-10-318-905-25

; Sequence 25, Application US/10318905

; Publication No. US20030152560A1

GENERAL INFORMATION:

; APPLICANT: Selden et al., Richard F.

; TITLE OF INVENTION: THERAPY FOR ALPHA-GALACTOSIDASE A

; DEFICIENCY

; NUMBER OF SEQUENCES: 28

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/318,905

; FILING DATE: 12-Dec-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/928,881

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Fraser, Janis K.

; REGISTRATION NUMBER: 34,819

; REFERENCE/DOCKET NUMBER: 07236/003001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617/542-5070

; TELEFAX: 617/542-8906

; INFORMATION FOR SEQ ID NO: 25:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1197 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 25:

US-10-318-905-25

Alignment Scores:
Pred. No.: 1.91e-06 Length: 1197
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x US-10-318-905-25 (1-1197)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTGTTCAGCTA 1158

RESULT 11

US-10-641-643-1001
; Sequence 1001, Application US/10641643
; Publication No. US20040077003A1
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; Susan G. Stuart
; Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL
; GENE EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESS: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/641,643
; FILING DATE: 14-Aug-2003
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1001:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1214 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g178245
; SEQUENCE DESCRIPTION: SEQ ID NO: 1001 :

Alignment Scores:
Pred. No.: 1.94e-06 Length: 1214
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-10-602-220-29 (1-18) x US-10-641-643-1001 (1-1214)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
Db 1122 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTGTTCAGCTA 1175

RESULT 12

US-09-993-059-15
; Sequence 15, Application US/09993059
; Publication No. US2002008024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-09-993-059-15

Alignment Scores:
Pred. No.: 2.02e-06 Length: 1254
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-993-059-15 (1-1254)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
Db 1198 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTGTTCAGCTA 1251

RESULT 13

US-10-103-327-15
; Sequence 15, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-10-103-327-15

Alignment Scores:
Pred. No.: 2.02e-06 Length: 1254
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-10-602-220-29 (1-18) x US-10-103-327-15 (1-1254)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
|||||
Db 1198 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACCTGTTTTCGTCAGCTA 1251

RESULT 14

US-10-602-219-15
; Sequence 15, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-219-15

Alignment Scores:
Pred. No.: 2,02e-06 Length: 1254
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-10-602-220-29 (1-18) x US-10-602-219-15 (1-1254)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
|||||
Db 1198 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACCTGTTTTCGTCAGCTA 1251

RESULT 15

US-10-602-220-15
; Sequence 15, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation

; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP07B
; CURRENT APPLICATION NUMBER: US/10/602,220
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-220-15

Alignment Scores:
Pred. No.: 2,02e-06 Length: 1254
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-10-602-220-29 (1-18) x US-10-602-220-15 (1-1254)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
|||||
Db 1198 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACCTGTTTTCGTCAGCTA 1251

Search completed: January 1, 2006, 09:34:34
Job time : 66.4276 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2006, 05:47:17 ; Search time 17.5034 Seconds

(without alignments)
539.100 Million cell updates/sec

Title: US-10-602-220-29

Perfect score: 89

Sequence: 1 TSLRSHINPTGTVLLQL 18

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4172979 seqs, 262114271 residues

Total number of hits satisfying chosen parameters: 8345958

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US10602220/runat_30122005_140602_5569/app_query.fasta 1.782
-DB=Published Applications NA.New -QPMT=fastcap -SUFFIX=p2n.rnpbn -MINMATCH=0.1
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10602220@cgn 1.1 233 @runat_30122005_140602_5569
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORE=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA.New.*

1: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq.*
2: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq.*
4: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq.*
5: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq.*
6: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq.*
7: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq.*
8: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq.*
9: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq.*
10: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	100.0	1355	US-10-981-267-26	Sequence 26, Appl
2	89	100.0	1547	US-10-981-267-25	Sequence 25, Appl
3	48	53.9	1716	US-10-750-185-42165	Sequence 42165, A
C 4	47	52.8	3321	US-10-750-185-61475	Sequence 61475, A
5	44	49.4	1380	US-10-750-185-56674	Sequence 56674, A
C 6	44	49.4	1794	US-11-074-176-257	Sequence 257, App
7	43	48.3	1587	US-10-131-826A-465	Sequence 465, App
8	43	48.3	1710	US-10-750-185-24794	Sequence 24794, A

C	9	43	48.3	1710	6	US-10-750-185-50667	Sequence 50667, A
	10	43	48.3	1861	6	US-10-750-185-43369	Sequence 43369, A
	11	43	48.3	100001	6	US-10-944-272-1	Sequence 1, Appl
	12	43	48.3	207835	7	US-11-121-086-39	Sequence 39, Appl
	13	43	48.3	207835	7	US-11-121-086-40	Sequence 40, Appl
	14	42	47.2	647	6	US-10-816-768-99	Sequence 99, Appl
	15	42	47.2	1916	6	US-10-750-185-38650	Sequence 38650, A
C	16	42	47.2	2217	6	US-10-750-185-40730	Sequence 40730, A
	17	42	47.2	2462	6	US-10-750-185-44108	Sequence 44108, A
	18	42	47.2	6855	6	US-10-750-185-32086	Sequence 32086, A
C	19	42	47.2	10129	7	US-11-044-111-21	Sequence 21, Appl
	20	42	47.2	31100	6	US-10-829-826B-24	Sequence 24, Appl
C	21	42	47.2	49979	6	US-10-995-561-13443	Sequence 13443, A
	22	42	47.2	184000	7	US-11-121-086-37	Sequence 37, Appl
	23	41	46.1	25	7	US-11-121-849-527128	Sequence 527128, A
	24	41	46.1	201	6	US-10-995-561-8091	Sequence 8091, Ap
	25	41	46.1	201	6	US-10-995-561-8098	Sequence 8098, Ap
	26	41	46.1	201	6	US-10-995-561-8132	Sequence 8132, Ap
	27	41	46.1	201	6	US-10-995-561-8172	Sequence 8172, Ap
	28	41	46.1	201	6	US-10-995-561-8179	Sequence 8179, Ap
	29	41	46.1	201	6	US-10-995-561-37193	Sequence 37193, A
C	30	41	46.1	201	6	US-10-995-561-41124	Sequence 41124, A
	31	41	46.1	201	6	US-10-995-561-73965	Sequence 73965, A
	32	41	46.1	699	6	US-10-750-185-25901	Sequence 25901, A
C	33	41	46.1	826	6	US-10-750-185-38506	Sequence 38506, A
	34	41	46.1	1156	6	US-10-750-185-45065	Sequence 45065, A
C	35	41	46.1	1201	6	US-10-750-185-27310	Sequence 27310, A
C	36	41	46.1	1883	6	US-10-750-185-47129	Sequence 47129, A
	37	41	46.1	2606	6	US-10-750-185-49768	Sequence 49768, A
	38	41	46.1	2703	7	US-11-191-072-1	Sequence 1, Appl
C	39	41	46.1	3885	6	US-10-793-626-3883	Sequence 3883, Ap
	40	41	46.1	4250	6	US-10-995-561-263	Sequence 263, App
	41	41	46.1	4432	6	US-10-995-561-265	Sequence 265, App
	42	41	46.1	4511	6	US-10-995-561-264	Sequence 264, App
	43	41	46.1	5700	6	US-10-513-786-7	Sequence 7, Appl
	44	41	46.1	5700	6	US-10-513-786-9	Sequence 9, Appl
	45	41	46.1	11115	6	US-10-513-786-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-10-981-267-26
; Sequence 26, Application US/10981267
; Publication No. US20050281805A1
; GENERAL INFORMATION:
; APPLICANT: LeBowitz, Jonathan H
; APPLICANT: Beverly, Stephen
; APPLICANT: Sly, William S.
; TITLE OF INVENTION: TARGETED THERAPEUTIC PROTEINS
; FILE REFERENCE: SYM-009CP
; CURRENT APPLICATION NUMBER: US/10/981,267
; CURRENT FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/384,452
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: US 60/516,990
; PRIOR FILING DATE: 2003-11-03
; PRIOR APPLICATION NUMBER: PCT/US03/17211
; PRIOR FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: US 10/272,531
; PRIOR FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US 60/445,734
; PRIOR FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: US 60/386,019
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: US 60/408,816
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 26
; LENGTH: 1355
; TYPE: DNA
; ORGANISM: Homo sapiens

US-10-981-267-26

Alignment Scores:
 Pred. No.: 3,99e-06 Length: 1355
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x US-10-981-267-26 (1-1355)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 Db 1245 ACTTCAAGGTGAAGTACATAATCCACAGGCACTGTTTTCCTTCAGCTA 1298

RESULT 2

US-10-981-267-25
 ; Sequence 25, Application US/10981267
 ; Publication No. US20050281805A1
 ; GENERAL INFORMATION:
 ; APPLICANT: LeBowitz, Jonathan H
 ; APPLICANT: Beverly, Stephen
 ; APPLICANT: Sly, William S.
 ; TITLE OF INVENTION: TARGETED THERAPEUTIC PROTEINS
 ; FILE REFERENCE: SYM-009CP
 ; CURRENT APPLICATION NUMBER: US/10/981,267
 ; PRIOR FILING DATE: 2004-11-03
 ; PRIOR APPLICATION NUMBER: US 60/384,452
 ; PRIOR FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: US 60/516,990
 ; PRIOR FILING DATE: 2003-11-03
 ; PRIOR APPLICATION NUMBER: PCT/US03/17211
 ; PRIOR FILING DATE: 2003-05-29
 ; PRIOR APPLICATION NUMBER: US 10/272,531
 ; PRIOR FILING DATE: 2002-10-16
 ; PRIOR APPLICATION NUMBER: US 60/445,734
 ; PRIOR FILING DATE: 2003-02-06
 ; PRIOR APPLICATION NUMBER: US 60/386,019
 ; PRIOR FILING DATE: 2002-06-05
 ; PRIOR APPLICATION NUMBER: US 60/408,816
 ; PRIOR FILING DATE: 2002-09-06
 ; NUMBER OF SEQ ID NOS: 27
 ; SOFTWARE: PatentIn version 3.3
 ; SEQ ID NO 25
 ; TYPE: DNA
 ; LENGTH: 1547
 ; FEATURE:
 ; OTHER INFORMATION: An exemplary GILT-tagged alpha-GAL A cassette sequence
 US-10-981-267-25

Alignment Scores:
 Pred. No.: 4.67e-06 Length: 1547
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x US-10-981-267-25 (1-1547)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 Db 1248 ACTTCAAGGTGAAGTACATAATCCACAGGCACTGTTTTCCTTCAGCTA 1301

RESULT 3

US-10-750-185-42165
 ; Sequence 42165, Application US/10750185
 ; Publication No. US20050260603A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MMI GENOMICS, INC.
 ; APPLICANT: Denise, Sue K.

; APPLICANT: KERR, Richard
 ; APPLICANT: ROSENFELD, David
 ; APPLICANT: HOLM, Tom
 ; APPLICANT: BATES, Stephen
 ; APPLICANT: FANTIN, Dennis
 ; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
 ; FILE REFERENCE: MM1100-2
 ; CURRENT APPLICATION NUMBER: US/10/750,185
 ; CURRENT FILING DATE: 2003-12-31
 ; PRIOR APPLICATION NUMBER: US 60/437,482
 ; PRIOR FILING DATE: 2002-12-31
 ; NUMBER OF SEQ ID NOS: 64922
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 42165
 ; LENGTH: 1716
 ; TYPE: DNA
 ; ORGANISM: Bovine 19866881152718
 ; US-10-750-185-42165

US-10-602-220-29 (1-18) x US-10-750-185-42165 (1-1716)

Qy 4 LeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 Db 1145 CTTCAACAGAACTTGAACTTACTTGAACCTTGATCTTAAGACTT 1189

RESULT 4

US-10-750-185-61475/c
 ; Sequence 61475, Application US/10750185
 ; Publication No. US20050260603A1
 ; GENERAL INFORMATION:
 ; APPLICANT: MMI GENOMICS, INC.
 ; APPLICANT: Denise, Sue K.
 ; APPLICANT: KERR, Richard
 ; APPLICANT: ROSENFELD, David
 ; APPLICANT: HOLM, Tom
 ; APPLICANT: BATES, Stephen
 ; APPLICANT: FANTIN, Dennis
 ; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
 ; FILE REFERENCE: MM1100-2
 ; CURRENT APPLICATION NUMBER: US/10/750,185
 ; CURRENT FILING DATE: 2003-12-31
 ; PRIOR APPLICATION NUMBER: US 60/437,482
 ; PRIOR FILING DATE: 2002-12-31
 ; NUMBER OF SEQ ID NOS: 64922
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 61475
 ; LENGTH: 3321
 ; TYPE: DNA
 ; ORGANISM: Bovine 19866880527707
 ; US-10-750-185-61475

Alignment Scores:
 Pred. No.: 546 Length: 3321
 Score: 47.00 Matches: 9
 Percent Similarity: 68.75% Conservative: 2
 Best Local Similarity: 56.25% Mismatches: 5
 Query Match: 52.81% Indels: 0
 DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x US-10-750-185-61475 (1-3321)
 Qy 3 ArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 Db 772 CGGCTCCATTGCCATGAGGCGCTCGTGGTCAGTCTTGGCTCTC 725

[illegible]

Query Match: 48.31% Indels: 0
DB: 6 Gaps: 0
US-10-602-220-29 (1-18) x US-10-131-826A-465 (1-1587)
Qy 3 ArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGln 17
Db 769 AGATGATGACTCACATCAACCTGGTGGGACAAAGGCTGCAG 813

RESULT 8

US-10-750-185-24794
; Sequence 24794, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24794
; LENGTH: 1710
; TYPE: DNA
; ORGANISM: Bovine 19866880168491
US-10-750-185-24794

Alignment Scores:
Pred. No.: 1.33e+03 Length: 1710
Score: 43.00 Matches: 8
Percent Similarity: 70.59% Conservative: 4
Best Local Similarity: 47.06% Mismatches: 5
Query Match: 48.31% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x US-10-750-185-24794 (1-1710)

Qy 2 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 175 AGCACTTAAGATCTCATGTACAGACATCCATACGATCATCTCCAAACTG 225

RESULT 9

US-10-750-185-50667/c
; Sequence 50667, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50667
; LENGTH: 1710
; TYPE: DNA
; ORGANISM: Bovine 19866880158521

US-10-750-185-50667

Alignment Scores:
Pred. No.: 1.33e+03 Length: 1710
Score: 43.00 Matches: 8
Percent Similarity: 78.57% Conservative: 3
Best Local Similarity: 57.14% Mismatches: 3
Query Match: 48.31% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x US-10-750-185-50667 (1-1710)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrVal 14
Db 707 AGCAGCGCTGAGCGCTCATATCATACCAAGTGGTTTAATC 666

RESULT 10

US-10-750-185-43369
; Sequence 43369, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43369
; LENGTH: 1861
; TYPE: DNA
; ORGANISM: Bovine 19866881226204
US-10-750-185-43369

Alignment Scores:

Pred. No.: 1.47e+03 Length: 1861
Score: 43.00 Matches: 8
Percent Similarity: 88.89% Conservative: 0
Best Local Similarity: 88.89% Mismatches: 1
Query Match: 48.31% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x US-10-750-185-43369 (1-1861)

Qy 4 LeuArgSerHisIleAsnProThrGly 12
Db 1561 CTCAGTCCACATCATCCCCCACCAGG 1587

RESULT 11

US-10-944-272-1
; Sequence 1, Application US/10944272
; Publication No. US20050272051A1
; GENERAL INFORMATION:
; APPLICANT: HELGADOTTIR et al.
; TITLE OF INVENTION: METHODS OF PREVENTING OR TREATING RECURRENCE OF MYOCARDIAL INFARCT
; FILE REFERENCE: 30847/2062-001
; CURRENT APPLICATION NUMBER: US/10/944,272
; CURRENT FILING DATE: 2004-09-17
; PRIOR APPLICATION NUMBER: 60/503,587
; PRIOR FILING DATE: 2003-09-17
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 100001
; TYPE: DNA


```
; ORGANISM: Homo sapiens
US-10-944-272-1

Alignment Scores:
Pred. No.: 1.72e+05      Length: 100001
Score: 43.00           Matches: 8
Percent Similarity: 76.92%  Conservatives: 2
Best Local Similarity: 61.54%  Mismatches: 3
Query Match: 48.31%      Indels: 0
DB: 6                  Gaps: 0

US-10-602-220-29 (1-18) x US-10-944-272-1 (1-100001)
Qy 5 ArgSerHisIleAsnProThrGlyThrValLeuGln 17
Db 4119 AGATCCCACTTGAGAACACCGGCACAGTGGTCTATCAG 4157

RESULT 12
US-11-121-086-39
; Sequence 39, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 39
; LENGTH: 207835
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-39

Alignment Scores:
Pred. No.: 4.07e+05      Length: 207835
Score: 43.00           Matches: 7
Percent Similarity: 69.23%  Conservatives: 2
Best Local Similarity: 53.85%  Mismatches: 4
Query Match: 48.31%      Indels: 0
DB: 7                  Gaps: 0

US-10-602-220-29 (1-18) x US-11-121-086-39 (1-207835)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThr 13
Db 9195 ACAGTGAGCTCCGCACTCATGTGGTCCATGTGGAACA 9233

RESULT 13
US-11-121-086-40
; Sequence 40, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 40
; LENGTH: 207835
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-40

Alignment Scores:
Pred. No.: 4.07e+05      Length: 207835
Score: 43.00           Matches: 7
Percent Similarity: 69.23%  Conservatives: 2
Best Local Similarity: 53.85%  Mismatches: 4
Query Match: 48.31%      Indels: 0
DB: 7                  Gaps: 0

US-10-602-220-29 (1-18) x US-11-121-086-40 (1-207835)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThr 13
Db 9195 ACAGTGAGCTCCGCACTCATGTGGTCCATGTGGAACA 9233

RESULT 14
US-10-816-768-99
; Sequence 99, Application US/10816768
; Publication No. US20050250936A1
; GENERAL INFORMATION:
; APPLICANT: Oppermann, Hermann
; APPLICANT: Tai, Mei-Sheng
; APPLICANT: McCartney, John
; TITLE OF INVENTION: Modified TGF-beta Superfamily Proteins
; FILE REFERENCE: STK-075
; CURRENT APPLICATION NUMBER: US/10/816,768
; CURRENT FILING DATE: 2004-04-02
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: PatentIn version 2.0
; SEQ ID NO 99
; LENGTH: 647
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: H2528
US-10-816-768-99

Alignment Scores:
Pred. No.: 630           Length: 647
Score: 42.00           Matches: 7
Percent Similarity: 71.43%  Conservatives: 3
Best Local Similarity: 50.00%  Mismatches: 4
Query Match: 47.19%      Indels: 0
DB: 6                  Gaps: 0

US-10-602-220-29 (1-18) x US-10-816-768-99 (1-647)
Qy 4 LeuArgSerHisIleAsnProThrGlyThrValLeuGln 17
Db 417 TTGGTTGCGACCTCGAGCCCAACCAACCATGCCATCATTCAG 458

RESULT 15
US-10-750-185-38650
; Sequence 38650, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38650
; LENGTH: 1916
; TYPE: DNA
```

Tue Jan 3 11:36:11 2006

! ORGANISM: Bovine 19866880881467
US-10-750-185-38650

Alignment Scores:
Pred. No.: 2.31e+03 Length: 1916
Score: 42.00 Matches: 9
Percent Similarity: 70.59% Conservative: 3
Best Local Similarity: 52.94% Mismatches: 5
Query Match: 47.19% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x US-10-750-185-38650 (1-1916)

Qy 2 SerArgLeuArgSerHisIleasnProThrGlyThrValLeuLeuGlnLeu 18
Db 955 TCTATGTTGGCTTCAGGTATAAGCCTAGTGGTTTCAGTATTTTTCAGATT 1005

Search completed: January 1, 2006, 09:56:14
Job time : 58.5034 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 23:47:52 ; Search time 0.910345 Seconds
(without alignments)
148.074 Million cell updates/sec

Title: US-10-602-220-29

Perfect score: 89

Sequence: 1 TSRLRSHNPTGTVLLQL 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 57103 seqs, 7488799 residues

Total number of hits satisfying chosen parameters: 57103

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA_New*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	47.2	102	6	US-10-816-768-84
2	42	47.2	102	6	US-10-816-768-85
3	42	47.2	102	6	US-10-816-768-86
4	42	47.2	102	6	US-10-816-768-87
5	42	47.2	102	6	US-10-816-768-88
6	42	47.2	203	6	US-10-816-768-100
7	41	46.1	101	7	US-11-191-072-14
8	41	46.1	102	6	US-10-816-768-83
9	41	46.1	119	7	US-11-191-072-12
10	41	46.1	119	7	US-11-191-072-13
11	41	46.1	501	7	US-11-191-072-2
12	41	46.1	1020	6	US-10-513-786-4
13	41	46.1	1041	6	US-10-995-561-780
14	41	46.1	1041	6	US-10-995-561-782
15	41	46.1	1097	6	US-10-995-561-781
16	41	46.1	1900	6	US-10-513-786-3
17	41	46.1	3704	6	US-10-513-786-1
18	38	42.7	218	6	US-10-467-657-1398
19	38	42.7	362	7	US-11-010-874-18
20	38	42.7	418	7	US-11-077-716-2
21	37	41.6	30	6	US-10-467-657-7384
22	37	41.6	402	6	US-10-821-234-1581
23	36	40.4	107	7	US-11-053-076-45
24	36	40.4	235	7	US-11-188-473-2
25	36	40.4	418	6	US-10-821-234-1331

ALIGNMENTS

RESULT 1

US-10-816-768-84
; Sequence 84, Application US/10816768
; Publication No. US20050250936A1
; GENERAL INFORMATION:
; APPLICANT: Oppermann, Hermann
; APPLICANT: Tai, Mei-Sheng
; APPLICANT: McCartney, John
; TITLE OF INVENTION: Modified TGF-beta Superfamily Proteins
; FILE REFERENCE: STK-075
; CURRENT APPLICATION NUMBER: US/10/816,768
; CURRENT FILING DATE: 2004-04-02
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: PatentIn version 2.0
; SEQ ID NO 84
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: CDMF-2/GDF-6
US-10-816-768-84

Query Match 47.2%; Score 42; DB 6; Length 102;
Best Local Similarity 50.0%; Pred. No. 0.83;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 LRSHINPTGTVLLQ 17
|||:|:|:
DB 38 LRSHLEPTNHAIQ 51

RESULT 2

US-10-816-768-85
; Sequence 85, Application US/10816768
; Publication No. US20050250936A1
; GENERAL INFORMATION:
; APPLICANT: Oppermann, Hermann
; APPLICANT: Tai, Mei-Sheng
; APPLICANT: McCartney, John
; TITLE OF INVENTION: Modified TGF-beta Superfamily Proteins
; FILE REFERENCE: STK-075
; CURRENT APPLICATION NUMBER: US/10/816,768
; CURRENT FILING DATE: 2004-04-02
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: PatentIn version 2.0
; SEQ ID NO 85
; LENGTH: 102
; TYPE: PRT

```
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: GDF-6
US-10-816-768-85

Query Match      47.2%; Score 42; DB 6; Length 102;
Best Local Similarity 50.0%; Pred. No. 0.83;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy      4 LRSHNPTGTVLLQ 17
      ||||: ||: ||
Db      38 LRSHLEPTNHAIQ 51

RESULT 3
US-10-816-768-86
; Sequence 86, Application US/10816768
; Publication No. US20050250936A1
; GENERAL INFORMATION:
; APPLICANT: Oppermann, Hermann
; APPLICANT: Tai, Mei-Sheng
; APPLICANT: McCartney, John
; TITLE OF INVENTION: Modified TGF-beta Superfamily Proteins
; FILE REFERENCE: STK-075
; CURRENT APPLICATION NUMBER: US/10/816,768
; CURRENT FILING DATE: 2004-04-02
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: PatentIn version 2.0
; SEQ ID NO 86
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: CDMF-2
US-10-816-768-86

Query Match      47.2%; Score 42; DB 6; Length 102;
Best Local Similarity 50.0%; Pred. No. 0.83;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy      4 LRSHNPTGTVLLQ 17
      ||||: ||: ||
Db      38 LRSHLEPTNHAIQ 51

RESULT 4
US-10-816-768-87
; Sequence 87, Application US/10816768
; Publication No. US20050250936A1
; GENERAL INFORMATION:
; APPLICANT: Oppermann, Hermann
; APPLICANT: Tai, Mei-Sheng
; APPLICANT: McCartney, John
; TITLE OF INVENTION: Modified TGF-beta Superfamily Proteins
; FILE REFERENCE: STK-075
; CURRENT APPLICATION NUMBER: US/10/816,768
; CURRENT FILING DATE: 2004-04-02
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: PatentIn version 2.0
; SEQ ID NO 87
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: GDF-7
US-10-816-768-87

Query Match      47.2%; Score 42; DB 6; Length 102;
Best Local Similarity 50.0%; Pred. No. 0.83;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy      4 LRSHNPTGTVLLQ 17
      ||||: ||: ||
Db      38 LRSHLEPTNHAIQ 51

RESULT 5
US-10-816-768-88
; Sequence 88, Application US/10816768
; Publication No. US20050250936A1
; GENERAL INFORMATION:
; APPLICANT: Oppermann, Hermann
; APPLICANT: Tai, Mei-Sheng
; APPLICANT: McCartney, John
; TITLE OF INVENTION: Modified TGF-beta Superfamily Proteins
; FILE REFERENCE: STK-075
; CURRENT APPLICATION NUMBER: US/10/816,768
; CURRENT FILING DATE: 2004-04-02
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: PatentIn version 2.0
; SEQ ID NO 88
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: CDMF-3 construct
US-10-816-768-88

Query Match      47.2%; Score 42; DB 6; Length 102;
Best Local Similarity 50.0%; Pred. No. 0.83;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy      4 LRSHNPTGTVLLQ 17
      ||||: ||: ||
Db      38 LRSHLEPTNHAIQ 51

RESULT 6
US-10-816-768-100
; Sequence 100, Application US/10816768
; Publication No. US20050250936A1
; GENERAL INFORMATION:
; APPLICANT: Oppermann, Hermann
; APPLICANT: Tai, Mei-Sheng
; APPLICANT: McCartney, John
; TITLE OF INVENTION: Modified TGF-beta Superfamily Proteins
; FILE REFERENCE: STK-075
; CURRENT APPLICATION NUMBER: US/10/816,768
; CURRENT FILING DATE: 2004-04-02
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: PatentIn version 2.0
; SEQ ID NO 100
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: H2528
US-10-816-768-100

Query Match      47.2%; Score 42; DB 6; Length 203;
Best Local Similarity 50.0%; Pred. No. 1.8;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy      4 LRSHNPTGTVLLQ 17
      ||||: ||: ||
Db      139 LRSHLEPTNHAIQ 152

RESULT 7
US-11-191-072-14
; Sequence 14, Application US/11191072
; Publication No. US20050282255A1
; GENERAL INFORMATION:
; APPLICANT: Hotten, Gertrud
; APPLICANT: Bechtold, Rolf
; APPLICANT: Pohl, Jens
```

; TITLE OF INVENTION: Monomeric Protein of the TGF-beta Family
; FILE REFERENCE: 2923-128
; CURRENT APPLICATION NUMBER: US/11/191,072
; PRIOR FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: US/10/048,458
; PRIOR FILING DATE: 2002-02-06
; PRIOR APPLICATION NUMBER: PCT/EP00/07600
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: EP 99115613.4
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 14
; PRIOR APPLICATION NUMBER: EP 99115613.4
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (66)..(66)
; OTHER INFORMATION: Xaa = any amino acid except cysteine
US-11-191-072-14

Query Match 46.1%; Score 41; DB 7; Length 101;
Best Local Similarity 50.0%; Pred. No. 1.2; Indels 4; Mismatches 3; Gaps 0;
Matches 7; Conservative 3; Indels 4; Mismatches 3; Gaps 0;

QY 4 LRSHINPTGTVLLQ 17
||||: || :
Db 38 LRSHLEPTNHAIVQ 51

RESULT 8
US-10-816-768-83
; Sequence 83, Application US/10816768
; Publication No. US20050250936A1
; GENERAL INFORMATION:
; APPLICANT: Oppermann, Hermann
; APPLICANT: Tai, Mei-Sheng
; APPLICANT: McCartney, John
; TITLE OF INVENTION: Modified TGF-beta Superfamily Proteins
; FILE REFERENCE: STK-075
; CURRENT APPLICATION NUMBER: US/10/816,768
; CURRENT FILING DATE: 2004-04-02
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: PatentIn version 2.0
; SEQ ID NO 83
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: CDMF-1/GDF-5
US-10-816-768-83

Query Match 46.1%; Score 41; DB 6; Length 102;
Best Local Similarity 50.0%; Pred. No. 1.2; Indels 4; Mismatches 3; Gaps 0;
Matches 7; Conservative 3; Indels 4; Mismatches 3; Gaps 0;

QY 4 LRSHINPTGTVLLQ 17
||||: || :
Db 38 LRSHLEPTNHAIVQ 51

RESULT 9
US-11-191-072-12
; Sequence 12, Application US/11191072
; Publication No. US20050282255A1
; GENERAL INFORMATION:
; APPLICANT: Hotten, Gertrud
; APPLICANT: Bechtold, Rolf
; APPLICANT: Pohl, Jens
; TITLE OF INVENTION: Monomeric Protein of the TGF-beta Family
; FILE REFERENCE: 2923-128
; CURRENT APPLICATION NUMBER: US/11/191,072

; CURRENT FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: US/10/048,458
; PRIOR FILING DATE: 2002-02-06
; PRIOR APPLICATION NUMBER: PCT/EP00/07600
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: EP 99115613.4
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 119
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: mutated recombinant human MP52
US-11-191-072-12

Query Match 46.1%; Score 41; DB 7; Length 119;
Best Local Similarity 50.0%; Pred. No. 1.5; Indels 4; Mismatches 3; Gaps 0;
Matches 7; Conservative 3; Indels 4; Mismatches 3; Gaps 0;

QY 4 LRSHINPTGTVLLQ 17
||||: || :
Db 55 LRSHLEPTNHAIVQ 68

RESULT 10
US-11-191-072-13
; Sequence 13, Application US/11191072
; Publication No. US20050282255A1
; GENERAL INFORMATION:
; APPLICANT: Hotten, Gertrud
; APPLICANT: Bechtold, Rolf
; APPLICANT: Pohl, Jens
; TITLE OF INVENTION: Monomeric Protein of the TGF-beta Family
; FILE REFERENCE: 2923-128
; CURRENT APPLICATION NUMBER: US/11/191,072
; CURRENT FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: US/10/048,458
; PRIOR FILING DATE: 2002-02-06
; PRIOR APPLICATION NUMBER: PCT/EP00/07600
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: EP 99115613.4
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 119
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: monomeric mutated recombinant human MP52
; NAME/KEY: MISC FEATURE
; LOCATION: (83)..(83)
; OTHER INFORMATION: Xaa = any amino acid except cysteine
US-11-191-072-13

Query Match 46.1%; Score 41; DB 7; Length 119;
Best Local Similarity 50.0%; Pred. No. 1.5; Indels 4; Mismatches 3; Gaps 0;
Matches 7; Conservative 3; Indels 4; Mismatches 3; Gaps 0;

QY 4 LRSHINPTGTVLLQ 17
||||: || :
Db 55 LRSHLEPTNHAIVQ 68

RESULT 11
US-11-191-072-2
; Sequence 2, Application US/11191072
; Publication No. US20050282255A1
; GENERAL INFORMATION:
; APPLICANT: Hotten, Gertrud

; APPLICANT: Bechtold, Rolf
; APPLICANT: Pohl, Jens
; TITLE OF INVENTION: Monomeric Protein of the TGF-beta Family
; FILE REFERENCE: 2923-128
; CURRENT APPLICATION NUMBER: US/11/191,072
; CURRENT FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: US/10/048,458
; PRIOR FILING DATE: 2002-02-06
; PRIOR APPLICATION NUMBER: PCT/EP00/07600
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: EP 99115613.4
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (465)..(465)
; OTHER INFORMATION: Xaa = any amino acid
US-11-191-072-2

Query Match 46.1%; Score 41; DB 7; Length 501;
Best Local Similarity 50.0%; Pred. No. 7.7; 4; Indels 4; Gaps 0;
Matches 7; Conservative 3; Mismatches 3

Qy 4 LRSHNPVTGTVLLQ 17
Db 437 LRSHLEPTNHAIVQ 450
||||: ||: ||: ||:

RESULT 12

US-10-513-786-4
; Sequence 4, Application US/10513786
; Publication No. US20050260589A1
; GENERAL INFORMATION:
; APPLICANT: Bio Control Institute Limited
; APPLICANT: OHTA, Michio
; APPLICANT: AGATA, Norio
; TITLE OF INVENTION: A cerulide synthetase, a gene thereof, and a detection method for
; TITLE OF INVENTION: cereulide.
; FILE REFERENCE: P0202401
; CURRENT APPLICATION NUMBER: US/10/513,786
; CURRENT FILING DATE: 2004-11-16
; PRIOR APPLICATION NUMBER: JP P2002-142398
; PRIOR FILING DATE: 2002-05-17
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1020
; TYPE: PRT
; ORGANISM: Bacillus cereus
US-10-513-786-4

Query Match 46.1%; Score 41; DB 6; Length 1020;
Best Local Similarity 57.9%; Pred. No. 17;
Matches 11; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

Qy 1 TSLRSHI--NPTGTVLLQ 17
Db 885 TLAURSHLDNPTVDVLLK 903
| ||||: ||: ||: ||:

RESULT 13

US-10-995-561-780
; Sequence 780, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF

; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 780
; LENGTH: 1041
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-780

Query Match 46.1%; Score 41; DB 6; Length 1041;
Best Local Similarity 60.0%; Pred. No. 18;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 5 RSHINPTGTV 14
Db 808 RQHLEPTGTI 817
| ||: ||||: ||:

RESULT 14

US-10-995-561-782
; Sequence 782, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 782
; LENGTH: 1041
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-782

Query Match 46.1%; Score 41; DB 6; Length 1041;
Best Local Similarity 60.0%; Pred. No. 18;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 5 RSHINPTGTV 14
Db 808 RQHLEPTGTI 817
| ||: ||||: ||:

RESULT 15

US-10-995-561-781
; Sequence 781, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 781
; LENGTH: 1097
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-781

Query Match 46.1%; Score 41; DB 6; Length 1097;
Best Local Similarity 60.0%; Pred. No. 19;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 5 RSHINPTGTV 14
| | : | | | :
Db 864 RQLEPTGII 873

Search completed: January 1, 2006, 00:29:53
Job time : 1.91034 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 23:36:26 ; Search time 1.86207 Seconds
(without alignments)
799.198 Million cell updates/sec

Title: US-10-602-220-29
Perfect score: 89
Sequence: 1 TSRLRSHINPTGTVLLQL 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.*
1: /cgn2_6/ptodata/1/1aa/5 COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/6 COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/H COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/PCTUS COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/RE COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	100.0	18	2	US-09-626-127-9
2	89	100.0	18	2	US-09-993-059-29
3	89	100.0	18	2	US-10-103-327-29
4	89	100.0	22	2	US-09-626-127-7
5	89	100.0	22	2	US-09-993-059-27
6	89	100.0	22	2	US-10-103-327-27
7	89	100.0	24	2	US-09-626-127-10
8	89	100.0	24	2	US-09-993-059-30
9	89	100.0	24	2	US-10-103-327-30
10	89	100.0	26	2	US-09-626-127-5
11	89	100.0	26	2	US-09-993-059-25
12	89	100.0	26	2	US-10-103-327-25
13	89	100.0	28	2	US-09-626-127-8
14	89	100.0	30	2	US-09-626-127-3
15	89	100.0	30	2	US-09-993-059-23
16	89	100.0	30	2	US-10-103-327-23
17	89	100.0	32	2	US-09-626-127-6
18	89	100.0	32	2	US-09-993-059-26
19	89	100.0	32	2	US-10-103-327-26
20	89	100.0	36	2	US-09-626-127-4
21	89	100.0	36	2	US-09-993-059-24
22	89	100.0	36	2	US-10-103-327-24
23	89	100.0	92	1	US-07-602-824A-9
24	89	100.0	92	1	US-07-983-451-9
25	89	100.0	92	1	US-08-261-577-12
26	89	100.0	386	2	US-09-176-666-11
27	89	100.0	387	2	US-09-176-666-10

```

28      89 100.0 388 2 US-09-176-666-9      Sequence 9, Appli
29      89 100.0 389 2 US-09-176-666-8      Sequence 8, Appli
30      89 100.0 390 2 US-09-176-666-7      Sequence 7, Appli
31      89 100.0 391 2 US-09-176-666-6      Sequence 6, Appli
32      89 100.0 392 2 US-09-176-666-5      Sequence 5, Appli
33      89 100.0 393 2 US-09-176-666-4      Sequence 4, Appli
34      89 100.0 394 2 US-09-176-666-3      Sequence 3, Appli
35      89 100.0 396 2 US-09-176-666-2      Sequence 2, Appli
36      89 100.0 398 2 US-08-928-881-26      Sequence 26, Appli
37      89 100.0 398 2 US-09-176-666-1      Sequence 1, Appli
38      89 100.0 398 2 US-09-543-921-26      Sequence 26, Appli
39      89 100.0 398 2 US-09-266-014-4       Sequence 4, Appli
40      89 100.0 398 2 US-09-491-759-26      Sequence 26, Appli
41      89 100.0 398 2 US-10-360-101-202      Sequence 202, App
42      89 100.0 417 2 US-09-993-059-16      Sequence 16, Appli
43      89 100.0 417 2 US-10-103-327-16      Sequence 16, Appli
44      89 100.0 421 2 US-09-993-059-12      Sequence 12, Appli
45      89 100.0 421 2 US-10-103-327-12

```

ALIGNMENTS

RESULT 1
US-09-626-127-9
; Sequence 9, Application US/09626127
; Patent No. 6846968
; GENERAL INFORMATION:
; APPLICANT: Garger, Stephen J.
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto
; TITLE OF INVENTION: Production of lysosomal enzymes in plants by transient expression
; FILE REFERENCE: 00801.0087.CPUS04
; CURRENT APPLICATION NUMBER: US/09/626,127
; CURRENT FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/003,737
; PRIOR FILING DATE: 1995-09-14
; PRIOR APPLICATION NUMBER: 07/170,771
; PRIOR FILING DATE: 1988-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 9
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-626-127-9
Query Match 100.0%; Score 89; DB 2; Length 18;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 TSRLRSHINPTGTVLLQL 18
    |||||
Db 1 TSRLRSHINPTGTVLLQL 18

```

RESULT 2
US-09-993-059-29
; Sequence 29, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0

```
; SEQ ID NO 29
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-09-993-059-29

Query Match      100.0%; Score 89; DB 2; Length 18;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 3
US-10-103-327-29
; Sequence 29, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-103-327-29

Query Match      100.0%; Score 89; DB 2; Length 18;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 4
US-09-626-127-7
; Sequence 7, Application US/09626127
; Patent No. 6846968
; GENERAL INFORMATION:
; APPLICANT: Garger, Stephen J.
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto
; TITLE OF INVENTION: Production of lysosomal enzymes in plants by transient expression
; FILE REFERENCE: 00801.0087.CPUS04
; CURRENT APPLICATION NUMBER: US/09/626,127
; CURRENT FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/003,737
; PRIOR FILING DATE: 1995-09-14
; PRIOR APPLICATION NUMBER: 07/170,771
; PRIOR FILING DATE: 1986-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-626-127-7

Query Match      100.0%; Score 89; DB 2; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.5e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 5
US-09-993-059-27
; Sequence 27, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-09-993-059-27

Query Match      100.0%; Score 89; DB 2; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.5e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 6
US-10-103-327-27
; Sequence 27, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-103-327-27

Query Match      100.0%; Score 89; DB 2; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.5e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 7
US-09-626-127-10
; Sequence 10, Application US/09626127
; Patent No. 6846968
; GENERAL INFORMATION:
; APPLICANT: Garger, Stephen J.
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto
; TITLE OF INVENTION: Production of lysosomal enzymes in plants by transient expression
; FILE REFERENCE: 00801.0087.CPUS04
; CURRENT APPLICATION NUMBER: US/09/626,127
; CURRENT FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/003,737
; PRIOR FILING DATE: 1995-09-14
; PRIOR APPLICATION NUMBER: 07/170,771
; PRIOR FILING DATE: 1986-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-626-127-10
```

; Sequence 10, Application US/09626127
; Patent No. 6846968
; GENERAL INFORMATION:
; APPLICANT: Garger, Stephen J.
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 00801.0087.CPUS04
; CURRENT APPLICATION NUMBER: US/09/626,127
; CURRENT FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/003,737
; PRIOR FILING DATE: 1995-09-14
; PRIOR APPLICATION NUMBER: 07/170,771
; PRIOR FILING DATE: 1988-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-626-127-10

Query Match 100.0%; Score 89; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSLRSHINPTGVLLQL 18
Db 1 TSLRSHINPTGVLLQL 18
|||||

RESULT 8

US-09-993-059-30
; Sequence 30, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-09-993-059-30

Query Match 100.0%; Score 89; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSLRSHINPTGVLLQL 18
Db 1 TSLRSHINPTGVLLQL 18
|||||

RESULT 9

US-10-103-327-30
; Sequence 30, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION

; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SEQ ID NO 30
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-103-327-30

Query Match 100.0%; Score 89; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSLRSHINPTGVLLQL 18
Db 1 TSLRSHINPTGVLLQL 18
|||||

RESULT 10

US-09-626-127-5
; Sequence 5, Application US/09626127
; Patent No. 6846968
; GENERAL INFORMATION:
; APPLICANT: Garger, Stephen J.
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 00801.0087.CPUS04
; CURRENT APPLICATION NUMBER: US/09/626,127
; CURRENT FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/003,737
; PRIOR FILING DATE: 1995-09-14
; PRIOR APPLICATION NUMBER: 07/170,771
; PRIOR FILING DATE: 1988-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-626-127-5

Query Match 100.0%; Score 89; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 3e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSLRSHINPTGVLLQL 18
Db 1 TSLRSHINPTGVLLQL 18
|||||

RESULT 11

US-09-993-059-25
; Sequence 25, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25

; LENGTH: 26
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-09-993-059-25

Query Match 100.0%; Score 89; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 3e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSLRSHINPTGTVLLQL 18
|||||
Db 1 TSLRSHINPTGTVLLQL 18
|||||

RESULT 12
US-10-103-327-25
; Sequence 25, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-103-327-25

Query Match 100.0%; Score 89; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 3e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSLRSHINPTGTVLLQL 18
|||||
Db 1 TSLRSHINPTGTVLLQL 18
|||||

RESULT 13
US-09-626-127-8
; Sequence 8, Application US/09626127
; Patent No. 6846968
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen J.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS04
; CURRENT APPLICATION NUMBER: US/09/626,127
; CURRENT FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/003,737
; PRIOR FILING DATE: 1995-09-14
; PRIOR APPLICATION NUMBER: 07/170,771
; PRIOR FILING DATE: 1988-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-626-127-8

Query Match 100.0%; Score 89; DB 2; Length 28;

Best Local Similarity 100.0%; Pred. No. 3.3e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSLRSHINPTGTVLLQL 18
|||||
Db 1 TSLRSHINPTGTVLLQL 18
|||||

RESULT 14
US-09-626-127-3
; Sequence 3, Application US/09626127
; Patent No. 6846968
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen J.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS04
; CURRENT APPLICATION NUMBER: US/09/626,127
; CURRENT FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/003,737
; PRIOR FILING DATE: 1995-09-14
; PRIOR APPLICATION NUMBER: 07/170,771
; PRIOR FILING DATE: 1988-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-626-127-3

Query Match 100.0%; Score 89; DB 2; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.6e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSLRSHINPTGTVLLQL 18
|||||
Db 1 TSLRSHINPTGTVLLQL 18
|||||

RESULT 15
US-09-993-059-23
; Sequence 23, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-09-993-059-23

Query Match 100.0%; Score 89; DB 2; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.6e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSLRSHINPTGTVLLQL 18
|||||
Db 1 TSLRSHINPTGTVLLQL 18
|||||

Search completed: December 31, 2005, 23:48:33
Job time : 1.86207 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 23:47:17 ; Search time 5.29655 Seconds
(without alignments)
1419.967 Million cell updates/sec

Title: US-10-602-220-29

Perfect score: 89

Sequence: 1 TSRLRSHINPTGTVLLQL 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA_Main:
1: /cgm2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgm2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgm2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgm2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgm2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgm2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	100.0	18	3	US-09-993-059-29
2	89	100.0	18	4	US-10-103-327-29
3	89	100.0	18	4	US-10-602-219-29
4	89	100.0	18	4	US-10-602-220-29
5	89	100.0	18	4	US-10-684-300-9
6	89	100.0	18	4	US-10-684-349-9
7	89	100.0	18	5	US-10-851-388-29
8	89	100.0	18	5	US-10-984-389-29
9	89	100.0	22	3	US-09-993-059-27
10	89	100.0	22	4	US-10-103-327-27
11	89	100.0	22	4	US-10-602-219-27
12	89	100.0	22	4	US-10-602-220-27
13	89	100.0	22	4	US-10-684-300-7
14	89	100.0	22	4	US-10-684-349-7
15	89	100.0	22	5	US-10-851-388-27
16	89	100.0	22	5	US-10-984-389-27
17	89	100.0	24	3	US-09-993-059-30
18	89	100.0	24	4	US-10-103-327-30
19	89	100.0	24	4	US-10-602-219-30
20	89	100.0	24	4	US-10-602-220-30
21	89	100.0	24	4	US-10-684-300-10
22	89	100.0	24	4	US-10-684-349-10
23	89	100.0	24	5	US-10-851-388-30
24	89	100.0	24	5	US-10-984-389-30
25	89	100.0	26	3	US-09-993-059-25
26	89	100.0	26	4	US-10-103-327-25
27	89	100.0	26	4	US-10-602-219-25

28	89	100.0	26	4	US-10-602-220-25
29	89	100.0	26	4	US-10-684-300-5
30	89	100.0	26	4	US-10-684-349-5
31	89	100.0	26	5	US-10-851-388-25
32	89	100.0	26	5	US-10-984-389-25
33	89	100.0	28	4	US-10-684-300-8
34	89	100.0	28	4	US-10-684-349-8
35	89	100.0	30	3	US-09-993-059-23
36	89	100.0	30	4	US-10-103-327-23
37	89	100.0	30	4	US-10-602-219-23
38	89	100.0	30	4	US-10-602-220-23
39	89	100.0	30	4	US-10-684-300-3
40	89	100.0	30	4	US-10-684-349-3
41	89	100.0	30	5	US-10-851-388-23
42	89	100.0	30	5	US-10-984-389-23
43	89	100.0	32	3	US-09-993-059-26
44	89	100.0	32	4	US-10-103-327-26
45	89	100.0	32	4	US-10-602-219-26

ALIGNMENTS

RESULT 1
US-09-993-059-29
; Sequence 29, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-09-993-059-29

Query Match 100.0%; Score 89; DB 3; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
|||
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 2
US-10-103-327-29
; Sequence 29, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 18
; TYPE: PRT

```
; ORGANISM: Tobacco mosaic virus
US-10-103-327-29

Query Match      100.0%; Score 89; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 3
US-10-602-219-29
; Sequence 29, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1999-05-21
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-602-219-29

Query Match      100.0%; Score 89; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 4
US-10-602-220-29
; Sequence 29, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Pogue, Gregory P.
```

```
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP07B
; CURRENT APPLICATION NUMBER: US/10/602,220
; CURRENT FILING DATE: 2003-06-23
; PRIOR FILING DATE: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-602-220-29

Query Match      100.0%; Score 89; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 5
US-10-684-300-9
; Sequence 9, Application US/10684300
; Publication No. US20040064855A1
; GENERAL INFORMATION:
; APPLICANT: Garger, Stephen J.
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto
; TITLE OF INVENTION: Production of lysosomal enzymes in plants by transient expression
; FILE REFERENCE: 00801.0087.CPUS04
; CURRENT APPLICATION NUMBER: US/10/684,300
; CURRENT FILING DATE: 2003-10-09
; PRIOR FILING DATE: 09/626,127
; PRIOR APPLICATION NUMBER: 2000-07-26
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/003,737
; PRIOR FILING DATE: 1995-09-14
; PRIOR APPLICATION NUMBER: 07/170,771
; PRIOR FILING DATE: 1988-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-684-300-9

Query Match      100.0%; Score 89; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



```
Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 6
US-10-684-349-9
; Sequence 9, Application US/10684349
; Publication No. US20040093646A1
; GENERAL INFORMATION:
; APPLICANT: Garger, Stephen J.
; APPLICANT: Turpen, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: Production of lysosomal enzymes in plants by transient expression
; FILE REFERENCE: 00801.0087.CPUS04
; CURRENT APPLICATION NUMBER: US/10/684,349
; CURRENT FILING DATE: 2003-10-09
; PRIOR FILING DATE: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR FILING DATE: 1995-09-14
; PRIOR FILING DATE: 07/170,771
; PRIOR FILING DATE: 1988-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-684-349-9

Query Match 100.0%; Score 89; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 7
US-10-851-388-29
; Sequence 29, Application US/10851388
; Publication No. US20040234516A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851,388
; CURRENT FILING DATE: 2004-05-21
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-851-388-29

Query Match 100.0%; Score 89; DB 5; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 8
US-10-984-389-29
; Sequence 29, Application US/10984389
; Publication No. US20050125859A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/984,389
; CURRENT FILING DATE: 2004-11-08
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-984-389-29

Query Match 100.0%; Score 89; DB 5; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 9
US-09-993-059-27
; Sequence 27, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-09-993-059-27

Query Match 100.0%; Score 89; DB 3; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 10
US-10-103-327-27
; Sequence 27, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
```

```
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-103-327-27
```

```
Query Match 100.0%; Score 89; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 TSRLRSHINPTGVLLQL 18
    |||||
Db 1 TSRLRSHINPTGVLLQL 18
    |||||
```

```
RESULT 11
US-10-602-219-27
; Sequence 27, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Poque, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
```

```
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 27
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-602-219-27

Query Match 100.0%; Score 89; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 TSRLRSHINPTGVLLQL 18
    |||||
Db 1 TSRLRSHINPTGVLLQL 18
    |||||
```

```
RESULT 12
US-10-602-220-27
; Sequence 27, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Poque, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP07B
; CURRENT APPLICATION NUMBER: US/10/602,220
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 27
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-602-220-27
```

```
Query Match 100.0%; Score 89; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 TSRLRSHINPTGVLLQL 18
    |||||
Db 1 TSRLRSHINPTGVLLQL 18
    |||||
```

```
RESULT 13
US-10-684-300-7
; Sequence 7, Application US/10684300
; Publication No. US20040064855A1
; GENERAL INFORMATION:
; APPLICANT: Garger, Stephen J.
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto
; TITLE OF INVENTION: Production of lysosomal enzymes in plants by transient expression
; FILE REFERENCE: 00801.0087.CFUS04
; CURRENT APPLICATION NUMBER: US/10/684,300
; CURRENT FILING DATE: 2003-10-09
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/003,737
; PRIOR FILING DATE: 1995-09-14
; PRIOR APPLICATION NUMBER: 07/170,771
; PRIOR FILING DATE: 1988-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 7
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-684-300-7

Query Match      100.0%; Score 89; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
   |||||
Db 1 TSRLRSHINPTGTVLLQL 18
   |||||

RESULT 14
US-10-684-349-7
; Sequence 7, Application US/10684349
; Publication No. US20040093646A1
; GENERAL INFORMATION:
; APPLICANT: Garger, Stephen J.
; APPLICANT: Turpen, Thomas H.
; APPLICANT: KUMAGAI, Monto
; TITLE OF INVENTION: Production of lysosomal enzymes in plants by transient expression
; FILE REFERENCE: 00801.0087.CPUS04
; CURRENT APPLICATION NUMBER: US/10/684,349
; CURRENT FILING DATE: 2003-10-09
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/003,737
; PRIOR FILING DATE: 1995-09-14
; PRIOR APPLICATION NUMBER: 07/170,771
; PRIOR FILING DATE: 1988-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-684-349-7

Query Match      100.0%; Score 89; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
   |||||
Db 1 TSRLRSHINPTGTVLLQL 18
   |||||

RESULT 15
US-10-851-388-27
; Sequence 27, Application US/10851388
; Publication No. US20040234516A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851.388
; CURRENT FILING DATE: 2004-05-21
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-851-388-27
```

```
Query Match      100.0%; Score 89; DB 5; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
   |||||
Db 1 TSRLRSHINPTGTVLLQL 18
   |||||

Search completed: January 1, 2006, 00:29:25
Job time : 6.29655 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 23:15:26 ; Search time 6.24828 Seconds
(without alignments)
2032.482 Million cell updates/sec

Title: US-10-602-220-29
Perfect score: 89
Sequence: 1 TSRLRSHINPTGTVLLQL 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	100.0	403	Q6LER7 HUMAN	Q6LER7 homo sapien
2	89	100.0	429	1 AGAL HUMAN	P06280 homo sapien
3	89	100.0	429	Q53HF3 HUMAN	Q53HF3 homo sapien
4	89	100.0	429	Q53Y83 HUMAN	Q53Y83 homo sapien
5	52	58.4	809	Q92Z25 RHIME	Q92Z25 rhizobium m
6	50	56.2	419	1 AGAL MOUSE	P51569 mus musculus
7	50	56.2	421	Q8BGZ6 MOUSE	Q8BGZ6 m mus muscu
8	47	52.8	206	Q5ND37 MOUSE	Q4tax9 mus musculus
9	47	52.8	312	Q4TAX9 TETNG	Q4tax9 tetraodon n
10	47	52.8	417	1 PRDF MOUSE	P97298 mus musculus
11	47	52.8	417	Q5ND38 MOUSE	Q5nd38 mus musculus
12	46	51.7	208	Q89KL6 BRAJA	Q89KL6 bradyrhizob
13	46	51.7	548	Q7SB68 NEUCR	Q7sb68 neurospora
14	46	51.7	609	Q22680 CAEEL	Q22680 caenorhabdi
15	45	50.6	213	Q7NXXD1 CHRVO	Q7nxd1 chromobacte
16	45	50.6	219	Q5WXY9 LEGPL	Q5wxy9 legionella
17	45	50.6	219	Q5X615 LEGPA	Q5x615 legionella
18	45	50.6	219	Q5XZ14 LEGPH	Q5xz14 legionella
19	45	50.6	221	Q82TY8 NITEU	Q82ty8 nitrosomona
20	45	50.6	240	Q8CMH3 STRA3	Q8cmh3 streptococc
21	45	50.6	306	Q7AMF9 NANEQ	Q7amf9 nanaorchaeu
22	45	50.6	512	1 SPEE OCEIH	Q8cv14 oceanobacil
23	45	50.6	542	1 RESB_BACSU	P3161 bacillus su
24	44	49.4	184	Q8FPT8 COREF	Q8fpt8 corynebacte
25	44	49.4	224	Q6N4M2 RHOPA	Q6n4m2 rhodopseudo
26	44	49.4	542	Q62TC8 BACLD	Q62tc8 bacillus li
27	44	49.4	543	Q65HX9 BACLD	Q65hx9 bacillus li
28	44	49.4	723	Q82670 CICAR	Q82670 cicer ariet
29	43	48.3	183	Q8TCQ3 HUMAN	Q8tcq3 homo sapien
30	43	48.3	196	1 Y1909 NEIMB	Q9jx82 neisseria m
31	43	48.3	286	1 AAC31_SALSP	P13246 salmonella

32	43	48.3	286	1 AAC32_SALSP	P0a255 salmonella
33	43	48.3	286	1 AAC22_ACIBA	P29807 acinetobact
34	43	48.3	286	1 AAC33_ENTCL	P0a256 enterobacte
35	43	48.3	286	2 Q52310_92ZZZ	Q52310 plasmid r.
36	43	48.3	286	2 Q5QJN0_SALTY	Q5qjn0 salmonella
37	43	48.3	286	2 Q7B8P8_PSEAE	Q7b8p8 pseudomonas
38	43	48.3	286	2 Q8GFQ5_CITFR	Q8gfc5 citrobacter
39	43	48.3	286	2 Q79NT9_ACIBA	Q79nt9 acinetobact
40	43	48.3	286	2 Q03634_ECOLI	Q03634 escherichia
41	43	48.3	286	2 Q6LDA3_ECOLI	Q6lda3 escherichia
42	43	48.3	296	1 Y608_TREPA	O83617 treponema p
43	43	48.3	338	2 Q9V2I0_PYRAB	Q9v2i0 pyrococcus
44	43	48.3	386	1 METK_FERTT	Q5nic7 francisella
45	43	48.3	403	2 Q4TSL8_9SPHN	Q4t818 erythrobact

ALIGNMENTS

RESULT 1
Q6LER7 HUMAN PRELIMINARY; PRT; 403 AA.
AC Q6LER7:
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE Alpha-galactosidase A (SC 3.2.1.22) (Fragment).
GN Name=alpha-Gala;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=lung;
RX MEDLINE=86259694; PubMed=3014515;
RA Bishop D.F., Calhoun D.H., Bernstein H.S., Hantzopoulos P., Quinn M.,
RA Desnick R.J.;
RT "Human alpha-galactosidase A: nucleotide sequence of a cDNA clone
encoding the mature enzyme.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:4859-4863(1986).
DR EMBL; D00039; BAA34059.1; -; mRNA.
DR SMR; Q6LER7; 6-396.
DR GO; GO:0004557; F:alpha-galactosidase activity; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR002241; Glyco_hydro_27.
DR InterPro; IPR000111; Glyco_hydro_GHD.
DR Pfam; PF02065; Melibiase; 1.
DR PRINTS; PR00740; GLYDRLASE27.
DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
KW Glycosidase; Hydrolase.
FT CHAIN 1
FT NON TER 6 403 alpha-galactosidase A subunit peptide.
SQ SEQUENCE 403 AA; 45804 MW; 2FE193205BEB8D1A CRC64;
Query Match 100.0%; Score 89; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 1.7e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSRLRSHINPTGTVLLQL 18
|||||
DB 374 TSRLRSHINPTGTVLLQL 391
RESULT 2
AGAL_HUMAN
ID AGAL_HUMAN STANDARD; PRT; 429 AA.
AC P06280;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)

DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alpha-galactosidase A precursor (EC 3.2.1.22) (Melibiase) (Alpha-D-
DE galactoside galactohydrolase) (Alpha-D-galactosidase A) (Agalsidase
DE alfa).
GN Name=GLA;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Lymphoblast;
RX MEDLINE=89263745; PubMed=2542896;
RA Kornreich R., Denick R.J., Bishop D.F.;
RT "Nucleotide sequence of the human alpha-galactosidase A gene.";
RL Nucleic Acids Res. 17:3301-3302(1989).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Fibroblast;
RX MEDLINE=87246603; PubMed=3036505;
RA Tsuji S., Martin B.M., Kaslow D.C., Migeon B.R., Choudary P.V.,
RA Stubblefield B.K., Mayor J.A., Murray G.J., Barranger J.A.,
RA Ginns E.I.;
RT "Signal sequence and DNA-mediated expression of human lysosomal alpha-
RT galactosidase A";
RL Eur. J. Biochem. 165:275-280(1987).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95352959; PubMed=7626884;
RA Oeltjen J.C., Liu X., Lu J., Allen R.C., Muzny D.M., Belmont J.W.,
RA Gibbs R.A.;
RT "Sixty-nine kilobases of contiguous human genomic sequence containing
RT the alpha-galactosidase A and Bruton's tyrosine kinase loci";
RL Mamm. Genome 6:334-338(1995).
RN [4]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15772651; DOI=10.1038/nature03440;
RA Ross M.T., Grafham D.V., Coffey A.J., Scherer S., McIay K., Muzny D.,
RA Platt M., Howell G.R., Burrows C., Bird C.P., Frankish A.,
RA Lovell F.L., Howe K.L., Ashurst J.L., Fulton R.S., Sudbrak R., Wen G.,
RA Jones M.C., Hultes M.E., Andrews T.D., Scott C.E., Searle S.,
RA Ramer J., Whittaker A., Deadman R., Carter N.P., Hunt S.E., Chen R.,
RA Cree A., Gunaratne P., Havlak P., Hodgson A., Metzker M.L.,
RA Richards S., Scott G., Steffen D., Sodergren E., Wheeler D.A.,
RA Worley K.C., Ainscough R., Ambrose K.D., Ansari-Lari M.A., Aradhy S.,
RA Ashwell R.I., Babbage A.K., Baggeley C.L., Ballabio A., Banerjee R.,
RA Barker G.E., Barlow K.F., Barrett I.P., Bates K.N., Beare D.M.,
RA Beasley H., Beasley O., Beck A., Bethel G., Blechschmidt K., Brady N.,
RA Bray-Allen S., Bridgeman A.M., Brown A.J., Brown M.J., Bonnin D.,
RA Bruford E.A., Buhay C., Burch P., Burford D., Burgess J., Burrill W.,
RA Burton J., Bye J.M., Carder C., Carrel L., Chako J., Chapman J.C.,
RA Chavez D., Chen E., Chen G., Chen Y., Chen Z., Chinault C.,
RA Ciccodicola A., Clark S.Y., Clarke G., Clee C.M., Clegg S.,
RA Clerc-Blankenburg K., Clifford K., Cobley V., Cole C.G., Conquer J.S.,
RA Corby N., Connor R.E., David R., Davies J., Davis C., Davis J.,
RA Delgado O., Deshazo D., Dhani P., Ding Y., Dinh H., Dodsworth S.,
RA Draper H., Dugan-Rocha S., Dunham A., Dunn M., Durbin K.J., Dutta I.,
RA Eades T., Ellwood M., Emery-Cohen A., Errington H., Evans K.L.,
RA Faulkner L., Francis F., Frankland J., Fraser A.E., Galoczky P.,
RA Gilbert J., Gill R., Gloeckner G., Gregory S.G., Gribble S.,
RA Griffiths C., Grocock R., Gu Y., Gwilliam R., Hamilton C., Hart E.A.,
RA Hawes A., Heath P.D., Heitmann K., Hennig S., Hernandez J.,
RA Hinzmann B., Ho S., Hoffs M., Howden P.J., Huckle E.J., Hume J.,
RA Hunt P.J., Hunt A.R., Inderwood J., Jacob L., Johnson D., Jones S.,
RA de Jong P.J., Joseph S.S., Keenan S., Kelly S., Kershaw J.K., Khan Z.,
RA Kioschis P., Klages S., Knights A.J., Kosiura A., Kovar-Smith C.,
RA Laird G.K., Langford C., Lawlor S., Leversha M., Lewis L., Liu W.,
RA Lloyd C., Lloyd D.M., Loueseged H., Loveland J.E., Lovell J.D.,
RA Lozadov R., Lu J., Lyne R., Ma J., Maheshwari M., Matthews L.H.,
RA McDowall J., McLaren S., McMurray A., Meidl P., Meitinger T.,
RA Milne S., Miner G., Mistry S.L., Morgan M., Morris S., Mueller I.,
RA Mullikin J.C., Nguyen N., Nordsiek G., Nyakatura G., O'dell C.N.,

RA Okwuonu G., Palmer S., Pandian R., Parker D., Parrish J.,
RA Pasternak S., Patel D.M., Pearce A.V., Pearson D.M., Pelan S.E.,
RA Perez L., Porter K.M., Ramsey Y., Reichwald K., Rhodes S.,
RA Ridler K.A., Schlessinger D., Schueler M.G., Sehra H.K.,
RA Shaw-Smith C., Shen H., Sheridan E.M., Shownkeen R., Skuce C.D.,
RA Smith K.L., Sotharan E.C., Steingrubber H.E., Steward C.A., Storey R.,
RA Swann R.M., Swabreck D., Tabor P.E., Taudien S., Taylor T.,
RA Teague B., Thomas K., Thorpe A., Timms K., Tracey A., Trevanion S.,
RA Tromans A.C., d'Urso M., Verdusco D., Villasana D., Waldron L.,
RA Wall M., Wang Q., Warren J., Warry G.L., Wei X., West A.,
RA Whitehead S.L., Whiteley M.N., Wilkinson J.E., Willey D.L.,
RA Williams G., Williams L., Williamson A., Williamson H., Wilming L.,
RA Woodmansey R.L., Wray P.W., Yen J., Zhang J., Zhou J., Zoghbi H.,
RA Zorilla S., Buck D., Reinhardt R., Poustka A., Rosenthal A.,
RA Lehrach H., Meindl A., Minx P.J., Hillier L.W., Willard H.F.,
RA Wilson R.K., Waterston R.H., Rice C.M., Vaudin M., Coulson A.,
RA Nelson D.B., Weinstein G., Sulston J.E., Durbin R., Hubbard T.,
RA Gibbs R.A., Beck S., Rogers J., Bentley D.R.;
RT "The DNA sequence of the human X chromosome.";
RL Nature 434:325-337(2005).
RN [5]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Uterus;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A.,
RA Fahey J., Helton E., Kettman M., Madan A.C., Shevchenko Y., Bouffard G.G.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Myers R.M.,
RA Buterfield Y.S.N., Kryzyski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP NUCLEOTIDE SEQUENCE OF 31-429, AND PARTIAL PROTEIN SEQUENCE.
RC TISSUE=Lung;
RX MEDLINE=86259694; PubMed=3014515;
RA Bishop D.F., Calhoun D.H., Bernstein H.S., Hantzopoulos P., Quinn M.,
RA Denick R.J.;
RT "Human alpha-galactosidase A: nucleotide sequence of a cDNA clone
RT encoding the mature enzyme.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:4859-4863(1986).
RN [7]
RP NUCLEOTIDE SEQUENCE OF 1-64.
RX MEDLINE=88112869; PubMed=2892762; DOI=10.1016/0378-1119(87)90374-X;
RA Quinn M., Hantzopoulos P., Fidanza V., Calhoun D.H.;
RT "A genomic clone containing the promoter for the gene encoding the
RT human lysosomal enzyme, alpha-galactosidase A";
RL Gene 58:177-188(1987).
RN [8]
RP NUCLEOTIDE SEQUENCE OF 1-64.
RX MEDLINE=88234528; PubMed=2836863;
RA Bishop D.F., Kornreich R., Denick R.J.;
RT "Structural organization of the human alpha-galactosidase A gene:
RT further evidence for the absence of a 3' untranslated region.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:3903-3907(1988).
RN [9]
RP RNA EDITING OF POSITION 396.
RX MEDLINE=95380278; PubMed=7503918;
RA Novo F.J., Kruszewski A., McDermot K.D., Goldspink G., Gorecki D.C.;
RT "Editing of human alpha-galactosidase RNA resulting in a pyrimidine to
RT purine conversion.";

RL Nucleic Acids Res. 23:2636-2640(1995).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (3.45 ANGSTROMS) OF 32-422 IN COMPLEX WITH
RC PRODUCT, HOMODIMERIZATION, AND N-GLYCOSYLATION.
RX PubMed=15003450; DOI=10.1016/j.jmb.2004.01.035;
RA Garman S.C., Garboczi D.N.;
RT "The molecular defect leading to Fabry disease: structure of human
RN alpha-galactosidase.";
RL J. Mol. Biol. 337:319-335(2004).
RN [11]
RP REVIEW ON PD VARIANTS.
RX MEDLINE=94258158; PubMed=7911050;
RA Eng C.M., Desnick R.J.;
RT "Molecular basis of Fabry disease: mutations and polymorphisms in the
RN human alpha-galactosidase A gene.";
RL Hum. Mutat. 3:103-111(1994).
RN [12]
RP VARIANT FD SER-40.
RX MEDLINE=90092580; PubMed=2152885; DOI=10.1016/0014-5793(90)80046-L;
RA Koide T., Ishiura M., Iwai K., Inoue M., Kaneda Y., Okada Y.,
RN Uchida T.;
RT "A case of Fabry's disease in a patient with no alpha-galactosidase A
RN activity caused by a single amino acid substitution of Pro-40 by
RN Ser.";
RL FEBS Lett. 259:353-356(1990).
RN [13]
RP VARIANT FD VAL-296.
RX MEDLINE=91101674; PubMed=1846223;
RA von Scheidt W., Eng C.M., Fitzmaurice T.F., Erdmann E., Hubner G.,
RN Olsen E.G.J., Christomanou H., Kandolf R., Bishop D.F., Desnick R.J.;
RT "An atypical variant of Fabry's disease with manifestations confined
RN to the myocardium.";
RL N. Engl. J. Med. 324:395-399(1991).
RN [14]
RP VARIANT FD GLN-301.
RX MEDLINE=91022721; PubMed=2171331;
RA Sakuraba H., Oshima A., Fukuhara Y., Shimmoto M., Nagao Y.,
RN Bishop D.F., Desnick R.J., Suzuki Y.;
RT "Identification of point mutations in the alpha-galactosidase A gene
RN in classical and atypical hemizygotes with Fabry disease.";
RL Am. J. Hum. Genet. 47:784-789(1990).
RN [15]
RP VARIANT FD TRP-356.
RX MEDLINE=89198098; PubMed=2539398;
RA Bernstein H.S., Bishop D.F., Astrin K.H., Kornreich R., Eng C.M.,
RN Sakuraba H., Desnick R.J.;
RT "Fabry disease: six gene rearrangements and an exonic point mutation
RN in the alpha-galactosidase gene.";
RL J. Clin. Invest. 83:1390-1399(1989).
RN [16]
Query Match 100.0%; Score 89; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.9e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 TSRLRSHINPTGTVLLQL 18
Db 400 TSRLRSHINPTGTVLLQL 417
RESULT 3
Q53HF3 HUMAN PRELIMINARY; PRT; 429 AA.
AC Q53HF3;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Galactosidase, alpha variant (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
NCBI_TaxID=9606;
Query Match 100.0%; Score 89; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.9e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 TSRLRSHINPTGTVLLQL 18
Db 400 TSRLRSHINPTGTVLLQL 417
RESULT 3
Q53HF3 HUMAN PRELIMINARY; PRT; 429 AA.
AC Q53HF3;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Galactosidase, alpha variant (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
NCBI_TaxID=9606;

RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Cerebellum;
RA Maruyama K., Sugano S.;
RT "Oligo-capping: a simple method to replace the cap structure of
RN eucaryotic mRNAs with oligoribonucleotides.";
RL Gene 138:171-174(1994).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Cerebellum;
RA Suzuki Y., Yoshitomo K., Maruyama K., Sugama A., Sugano S.;
RT "Construction and characterization of a full length-enriched and a 5'-
RN end-enriched cDNA library.";
RL Gene 200:149-156(1997).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Cerebellum;
RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,
RN Tanaka A., Yokoyama S.;
RT Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK222627; BAD96347.1; -; mRNA.
FT NON TER 1
SQ SEQUENCE 429 AA; 48752 MW; AFA828F5B22BED76 CRC64;
Query Match 100.0%; Score 89; DB 2; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.9e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 TSRLRSHINPTGTVLLQL 18
Db 400 TSRLRSHINPTGTVLLQL 417
RESULT 4
Q53Y83 HUMAN PRELIMINARY; PRT; 429 AA.
ID Q53Y83;
AC Q53Y83;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Galactosidase, alpha.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Kainine N., Chen X., Rolfs A., Halleck A., Hines L., Eisenstein S.,
RN Koundinya M., Raphael J., Moreira D., Kelley T., LaBaer J., Lin Y.,
RA Phelan M., Farmer A.;
RT "Cloning of human full-length CDSs in BD Creator(TM) System Donor
RN vector.";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BT006864; AAP35510.1; -; mRNA.
SQ SEQUENCE 429 AA; 48766 MW; 613F8FB21B107D7B CRC64;
Query Match 100.0%; Score 89; DB 2; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.9e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 TSRLRSHINPTGTVLLQL 18
Db 400 TSRLRSHINPTGTVLLQL 417
RESULT 5
Q92Z25 RHIME PRELIMINARY; PRT; 809 AA.
ID Q92Z25;
AC Q92Z25;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

```

DE Adenylate cyclase, putative.
GN OrderedLocusNames=RA0303; ORFNames=SMa0579;
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymA.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432; DOI=10.1073/pnas.161294798;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
DR EMBL; AF007222; XAK64961.1; -; Genomic_DNA.
DR F01; G95299; G95299.
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004383; F:guanylate cyclase activity; IEA.
DR GO; GO:0016823; F:lyase activity; IEA.
DR GO; GO:0000287; F:magnesium ion binding; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006171; F:cAMP biosynthesis; IEA.
DR GO; GO:0007242; F:intracellular signaling cascade; IEA.
DR InterPro; IPR001054; G:Cyclase.
DR InterPro; IPR003660; His_kin_HAMP.
DR Pfam; PF00211; Guanylate_cyc_1.
DR Pfam; PF00672; HAMP; 1.
DR SMART; SM00044; CYCC; 1.
DR SMART; SM00304; HAMP; 1.
DR PROSITE; PS0125; GUANYLATE_CYCLASES_2; 1.
DR PROSITE; PS00885; HAMP; 1.
KW Complete proteome; Plasmid.
SQ SEQUENCE 809 AA; 88948 MW; F650FFE4BD67BA21 CRC64;

Query Match 58.4%; Score 52; DB 2; Length 809;
Best Local Similarity 66.7%; Pred. No. 9.2;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 RLRSHPNTGTVLQ 17
DB 486 RLEAHRNPTGTLQ 500

RESULT 6
AGAL_MOUSE STANDARD; PRT; 419 AA.
AC P51569;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alpha-galactosidase A precursor (EC 3.2.1.22) (Melibiose) (Alpha-D-
DE galactoside galactohydrolase) (Alpha-D-galactosidase A).
GN Name=Gla; Synonyms=Ags;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Kidney;
RX MEDLINE=96125203; PubMed=85431175; DOI=10.1016/0378-1119(95)00592-7;
RA Ohshima T., Murray G.J., Negie J.W., Quirk J.M., Kraus M.H.,
RA Barton N.W., Brady R.O., Kulkarni A.B.;
RT "Structural organization and expression of the mouse gene encoding
RT alpha-galactosidase A."
RL Gene 166:277-280(1995).

```

```

RN NUCLEOTIDE SEQUENCE.
RP STRAIN=C129;
RX MEDLINE=95352959; PubMed=7626884;
RA Oeltjen J.C., Liu X., Lu J., Allen R.C., Muzny D.M., Belmont J.W.,
RA Gibbs R.A.;
RT "Sixty-nine kilobases of contiguous human genomic sequence containing
RT the alpha-galactosidase A and Bruton's tyrosine kinase loci."
RL Mamm. Genome 6:334-338(1995).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=96316016; PubMed=8733892; DOI=10.1006/bmme.1996.0020;
RA Gotlib R.W., Bishop D.F., Wang A.M., Zeidner K.M., Ioannou Y.I.,
RA Adler D.A., Distèche C.M., Desnick R.J.;
RT "The entire genomic sequence and cDNA expression of mouse alpha-
RT galactosidase A."
RL Biochem. Mol. Med. 57:139-148(1996).
RN [4]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC STRAIN=FVB/N; TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Pahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalusz D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing alpha-D-
CC galactose residues in alpha-D-galactosides, including galactose
CC oligosaccharides, galactomannans and galactohydrolase.
CC -!- SUBUNIT: Homodimer.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 27 family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; U34071; AAA96749.1; -; mRNA.
CC EMBL; L46651; AAA74453.1; -; Genomic DNA.
CC EMBL; U58105; ABA47244.1; -; Genomic DNA.
CC EMBL; U50716; AAC52584.1; -; mRNA.
CC EMBL; U50715; AAC52583.1; -; Genomic DNA.
CC EMBL; BC009021; AAH09021.1; -; mRNA.
CC PIR; JC4522; JC4522.
CC SMR; P51569; 32-418.
CC Ensembl; ENSMUSG00000031266; Mus musculus.
CC MGI; MGI:1347344; Gla.
CC GO; GO:0005615; C:extracellular space; TAS.
CC InterPro; IPR002241; Glyco_hydro_27.
CC InterPro; IPR000111; Glyco_hydro_GHD.
CC Pfam; PF02065; Melibiase; 1.
CC PRINTS; PR00740; GLHYDRASE27.
CC PRODOM; PD002572; Glyco_hydro_GHD; 1.
CC PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
KW Glycoprotein; Glycosidase; Hydrolase; Lysosome; Signal.
FT SIGNAL 1 31

```



```

FT CHAIN 32 419 Alpha-galactosidase A.
FT ACT SITE 170 Nucleophile (By similarity).
FT ACT SITE 231 Proton donor (By similarity).
FT SITE 203 Substrate binding (By similarity).
FT CARBOHYD 139 N-linked (GlcNAc...) (By similarity).
FT CARBOHYD 192 N-linked (GlcNAc...) (By similarity).
FT CARBOHYD 215 N-linked (GlcNAc...) (By similarity).
FT DISULFID 52 By similarity.
FT DISULFID 56 By similarity.
FT DISULFID 142 By similarity.
FT DISULFID 202 By similarity.
FT DISULFID 378 By similarity.
SQ SEQUENCE 419 AA; 47643 MW; BDSE6A99AC113613 CRC64;

Query Match 56.2%; Score 50; DB 1; Length 419;
Best Local Similarity 50.0%; Pred. No. 9.5;
Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 TSLRSHNPQTGLVQL 18
DB 400 TLTLRVNSPTGLFRL 417

RESULT 7
QBEG26 MOUSE
ID QBEG26 MOUSE PRELIMINARY; PRT; 421 AA.
AC QBEG26;
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched
DE library, clone:A430057F16 product:galactosidase, alpha, full insert
DE sequence (Mus musculus 2 days pregnant adult female ovary cDNA, RIKEN
DE full-length enriched library, clone:E330039P08 product:galactosidase,
DE alpha, full insert sequence).
CN Name=Gla;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Mateu Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staab F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guastincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazarelli J., Mombaerts P.,
RA Nordone P., King B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RN Nature 409:685-690(2001).
RN [3]

```

```

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Oeato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Guastincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawai J., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C.F., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Yasunishi A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multiplexed capillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hasegawa T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Oheato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK040081; BAC30508.1; -; mRNA.
DR EMBL; AK054547; BAC35819.1; -; mRNA.
DR HSSP; P06280; 1R46.

```

```

DR SMR: Q8BGZ6; 34-420.
DR MGI:1347344; Glia.
DR GO:GO:0005615; C:extracellular space; TAS.
DR InterPro: IPR002241; Glyco_hydro_27.
DR DR Pfam: PF02065; Melibiase; 1.
DR PRINTS: PR00740; GLHYDRLASE27.
DR PRODOM: PD002572; Glyco_hydro_GHD; 1.
DR PROSITE: PS00512; ALPHA GALACTOSIDASE; 1.
SQ SEQUENCE 421 AA; 47844 MW; 5F66772334014B6F CRC64;

Query Match 56.2%; Score 50; DB 2; Length 421;
Best Local Similarity 50.0%; Pred. No. 9.6;
Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 TSLRLSHINPTGTVLLQL 18
| : : : : :
Db 402 TLTLKTRVPSGTVLFLR 419

RESULT 8
Q5ND37 MOUSE PRELIMINARY; PRT; 206 AA.
AC Q5ND37;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Serine (Or cysteine) proteinase inhibitor, clade F, member 1
DE (Fragment).
GN Name=Serpinf1; ORFNames=RP23-384C18.1-002;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1];
RA Phillimore B.;
RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL591496; CAI35379.1; -; Genomic DNA.
DR GO:GO:0004867; F:serine-type endopeptidase inhibitor activity; IEA.
DR InterPro: IPR00215; Prot_inh_serpin.
DR Pfam: PF00079; Serpin; 2.
DR NON_TER 206 206
FT SEQUENCE 206 AA; 22551 MW; F24295B0571CCC5C CRC64;

Query Match 52.8%; Score 47; DB 2; Length 206;
Best Local Similarity 71.4%; Pred. No. 14;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 RLRSHPINPTGTVLL 16
| : : : : :
Db 66 RLRSASPTGNVLL 79

RESULT 9
Q4TAX9 TETNG PRELIMINARY; PRT; 312 AA.
AC Q4TAX9;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Chromosome undetermined SCAF7240, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG00004000001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontoides; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1];
RA Nishikawa Y., Shimizu S., Saga S.;

```

```

RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poullain J., De Bernardis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Landt V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -!- SIMILARITY: Belongs to the TGF-beta family.
DR EMBL; CAAE01007240; CAF89953.1; -; Genomic DNA.
DR InterPro: IPR001839; TGFb.
DR Pfam: PF00019; TGF_beta; 1.
DR PRODOM: PD000357; TGFb; 1.
DR SMART; SM00204; TGFb; 1.
DR PROSITE; PS00250; TGF_BETA_1; 1.
DR KW Growth factor.
FT NON_TER 1 1
FT NON_TER 312 312
SQ SEQUENCE 312 AA; 33461 MW; D9AABD81B0B9FD79 CRC64;

Query Match 52.8%; Score 47; DB 2; Length 312;
Best Local Similarity 53.3%; Pred. No. 22;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 RLRSHPINPTGTVLLQ 17
| : : : : :
Db 252 RLRSHPINPTGTVLLQ 266

RESULT 10
PEDF_MOUSE STANDARD; PRT; 417 AA.
ID PEDF_MOUSE STANDARD; PRT; 417 AA.
AC P97298; O70629; O88691;
DT 15-JUL-1998 (Rel. 36, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Pigment epithelium-derived factor precursor (PEDF) (Stromal cell-
DE derived factor 3) (SDF-3) (Caspin).
GN Name=Serpinf1; Synonym=Pedf, Sdf3;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1];
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=97092876; PubMed=8938438; DOI=10.1006/geno.1996.0560;
RA Shirozu M., Tada H., Tashiro K., Nakamura T., Lopez N.D., Nazarea M.,
RA Hamada T., Sato T., Nakano T., Honjo T.;
RT "Characterization of novel secreted and membrane proteins isolated by
RT the signal sequence trap method.";
RL Genomics 37:273-280(1996).
RN [2]
RP NUCLEOTIDE SEQUENCE. PROTEIN SEQUENCE OF 53-66; 252-261; 333-344 AND
RP 359-372. TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RC STRAIN=BAALB/c; TISSUE=Liver;
RX MEDLINE=98279032; PubMed=9614124; DOI=10.1074/jbc.273.24.15125;
RA Kozaki K., Miyaishi O., Koizumi Y., Yasui Y., Kashiwa A.,
RA Nishikawa Y., Shimizu S., Saga S.;

```

RT "Isolation, purification and characterization of a collagen-associated
 RL serpin, caspin, produced by murine colon adenocarcinoma cells.",
 J. Biol. Chem. 273:15125-15130(1998).
 [3]
 RN NUCLEOTIDE SEQUENCE, AND TISSUE SPECIFICITY.
 RP STRAIN=FVB/N; TISSUE=Liver;
 RX PubMed=5565647;
 RA Singh V.K., Chader G.J., Rodriguez I.R.;
 RT "Structural and comparative analysis of the mouse gene for pigment
 RL epithelium-derived factor (PEDF).";
 Mol. Vis. 4:7-7(1998).
 [4]
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=BA1B/c;
 RX Tombran-Tink J.;
 RA Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 [5]
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RP STRAIN=FVB/N; TISSUE=Liver;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S.S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RL and mouse cDNA sequences.";
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [6]
 RN FUNCTION.
 RP PubMed=12632345; DOI=10.1053/jpsu.2003.50104;
 RX Abramson L.P., Stellmach V., Doll J.A., Cornwell M., Arensman R.M.,
 RA Crawford S.E.;
 RT "Wilms' tumor growth is suppressed by antiangiogenic pigment
 RL epithelium-derived factor in a xenograft model.";
 J. Pediatr. Surg. 38:336-342(2003).
 CC -|- FUNCTION: Neurotrophic protein; induces extensive neuronal
 CC differentiation in retinoblastoma cells. Potent inhibitor of
 CC angiogenesis. As it does not undergo the S (stressed) to R
 CC (relaxed) conformational transition characteristic of active
 CC serpins, it exhibits no serine protease inhibitory activity.
 CC -|- SUBCELLULAR LOCATION: Secreted.
 CC -|- TISSUE SPECIFICITY: Highly expressed in the liver, gastric
 CC glandular mucosa and renal tubules. It is also expressed in the
 CC brain, heart, lung retina and testes.
 CC -|- DEVELOPMENTAL STAGE: First detected at 12.5 dpc in cartilage
 CC primordium, it is present in the osseous matrix of developing
 CC limbs, vertebrae, ribs and skull. At 16.5 dpc it is detected in
 CC bone matrix and smooth muscle, and at lower levels in connective
 CC tissue, bronchial epithelial cells, metanephron microtubules, and
 CC skin.
 CC -|- SIMILARITY: Belongs to the serpin family.
 CC
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC
 CC EMBL; D50460; BAA09051.1; -; mRNA.
 CC EMBL; AF036164; AAC69271.1; -; mRNA.

DR EMBL; D87975; BAA31978.1; -; mRNA.
 DR EMBL; AF017055; AAC05733.1; -; Genomic DNA.
 DR EMBL; AF017051; AAC05733.1; JOINED; Genomic DNA.
 DR EMBL; AF017052; AAC05733.1; JOINED; Genomic DNA.
 DR EMBL; AF017053; AAC05733.1; JOINED; Genomic DNA.
 DR EMBL; AF017054; AAC05733.1; JOINED; Genomic DNA.
 DR EMBL; AF017057; AAC05731.1; -; mRNA.
 DR EMBL; BC019852; AAH19852.1; -; mRNA.
 DR HSSP; P36955; 1MW.
 DR Ensemble; ENSMUSG0000000753; Mus musculus.
 DR MGI; MGI:108080; Serpinf1.
 DR GO; GO:0005635; C:extracellular space; IDA.
 DR GO; GO:0016525; P:negative regulation of angiogenesis; IDA.
 DR GO; GO:0050769; P:positive regulation of neurogenesis; ISS.
 DR InterPro; IPR000215; Prot_inh_serpin.
 DR PANTHER; PTHR11461; Prot_inh_serpin; 1.
 DR Pfam; PF00079; Serpin; 1.
 DR SMART; SM00093; SERPIN; 1.
 DR PROSITE; PS00284; SERPIN; 1.
 DR Direct protein sequencing; Glycoprotein; Serpin; Signal.
 KW SIGNAL 1 19
 FT CHAIN 20 417
 FT CARBOHYD 284 284
 FT CONFLICT 70 70
 FT CONFLICT 136 136
 FT CONFLICT 137 137
 FT CONFLICT 280 280
 FT CONFLICT 377 377
 SQ SEQUENCE 417 AA; 46234 MW; ECD360FEGAA74D25 CRC64;
 Query Match 52.8%; Score 47; DB 1; Length 417;
 Best Local Similarity 71.4%; Pred. No. 31;
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 Qy 3 RLRSHINPTGVLL 16
 Db 66 RLRSASPTGNVLL 79
 RESULT 11
 Q5ND38 MOUSE PRELIMINARY; PRT; 417 AA.
 ID Q5ND38_MOUSE PRELIMINARY; PRT; 417 AA.
 AC Q5ND38;
 DT 01-FEB-2005 (TrEMBLrel. 29, Created)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
 DE Serine (Or cysteine) proteinase inhibitor, clade F, member 1.
 GN Name=Serpinf1; ORFNames=RP23-384C18.1-001;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Phillimore B.;
 RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
 CC -|- SIMILARITY: Belongs to the serpin family.
 DR EMBL; AL591496; CA135378.1; -; Genomic DNA.
 DR GO; GO:0004867; F:serine-type endopeptidase inhibitor activity; IEA.
 KW Serpin.
 SQ SEQUENCE 417 AA; 46234 MW; ECD360FEGAA74D25 CRC64;
 Query Match 52.8%; Score 47; DB 2; Length 417;
 Best Local Similarity 71.4%; Pred. No. 31;
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 Qy 3 RLRSHINPTGVLL 16
 Db 66 RLRSASPTGNVLL 79
 RESULT 12

```

Q89KL6 BRAJA
ID Q89KL6_BRAJA PRELIMINARY; PRT; 208 AA.
AC Q89KL6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Transcriptional regulatory protein.
GN OrderedLocusNames=blr4889;
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Ideawa K., Iriiguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
CC -!- SIMILARITY: Contains 1 HTH LysR-type DNA-binding domain.
DR EMBL; BA000040; BAC50154.1; -; Genomic_DNA.
DR HSPF; Q9WCX7; IIXC.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR InterPro; IPR000847; HTH_LysR.
DR InterPro; IPR011991; Wing_hlx_DNA_bd.
DR Pfam; PF0126; HTH_1; 1.
DR PRINTS; PR00039; HTHLYSR.
DR PROSITE; PS0931; HTH_LYSR; 1.
KW Complete proteome; DNA-binding; Transcription;
KW Transcription regulation.
SQ SEQUENCE 208 AA; 22591 MW; 5C255F49DAC0D5DB CRC64;

Query Match 51.7%; Score 46; DB 2; Length 208;
Best Local Similarity 69.2%; Pred. No. 21;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 6 SHINPTGTVLLQ 18
| : : : : :
DB 85 SOLKPSGTVLLQL 97

RESULT 13
Q7SB68 NEUCR
ID Q7SB68_NEUCR PRELIMINARY; PRT; 548 AA.
AC Q7SB68;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Predicted protein.
GN Name=NCU06273.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Selitrenkoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamysseilis M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA Krystofova S., Rasmussen C., Metzberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmani S.A.,

```

```

RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Eboile D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nubaum C., Birren B.;
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0:0-0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC preliminary data.
CC EMBL/GenBank/DDAJ whole genome shotgun (WGS) entry which is
DR EMBL; AABX01000173; EAA33641.1; -; Genomic_DNA.
SQ SEQUENCE 548 AA; 61216 MW; 7E23FDFD00D83219 CRC64;

Query Match 51.7%; Score 46; DB 2; Length 548;
Best Local Similarity 52.9%; Pred. No. 63;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGTVLLQ 17
| : : : : :
DB 399 TAALRRYLNPPTTVRL 415

RESULT 14
Q22680 CAEBL
ID Q22680 CAEBL PRELIMINARY; PRT; 609 AA.
AC Q22680;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein T22C8.7.
GN ORFNames=T22C8.7;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peleoderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Bristol N2;
RX MEDLINE=93069613; PubMed=9851916;
RG The C. elegans sequencing consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z49071; CAA88879.1; -; Genomic_DNA.
DR PIR; T25120; T25120.
DR Ensembl; T22C8.7; Caenorhabditis elegans.
DR WormBase; WBGene00011928; T22C8.7.
DR WormPep; T22C8.7; CE02355.
DR InterPro; IPR001507; Endoglin/CD105.
DR Pfam; PF00100; Zona_pellucida; 1.
DR SMART; SM00241; ZP_1.
DR PROSITE; PS51034; ZP_2; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 609 AA; 68883 MW; 1DC9BA97CAFBCA2D CRC64;

Query Match 51.7%; Score 46; DB 2; Length 609;
Best Local Similarity 80.0%; Pred. No. 72;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LRSHINPTGT 13
| : : : : :
DB 43 IRLHINPTGT 52

RESULT 15
Q7NXD1 CHRVO
ID Q7NXD1_CHRVO PRELIMINARY; PRT; 213 AA.
AC Q7NXD1;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Probable integral membrane protein.
GN OrderedLocusNames=CV1696;
OS Chromobacterium violaceum.

```

OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Chromobacterium.
 OX NCBI_TaxID=536;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=ATCC 12472 / DSM 30191;
 RX MEDLINE=22882880; PubMed=14500782; DOI=10.1073/pnas.1832124100;
 RA Vasconcelos A.T.R., de Almeida D.F., Hungria M., Guimaraes C.T.,
 RA Antonio R.V., Almeida F.C., de Almeida L.G.P., de Almeida R.,
 RA Alves-Gomes J.A., Andrade E.M., Araripe J., de Araujo M.F.F.,
 RA Astolfi-Filho S., Azevedo V., Baptista A.J., Bataus L.A.M.,
 RA Batista J.S., Belo A., van den Berg C., Bogo M., Bonatto S.,
 RA Bordignon J., Brigido M.M., Brito C.A., Brocchi M., Burity H.A.,
 RA Camargo A.A., Cardoso D.D.P., Carneiro N.P., Carriaro D.M.,
 RA Carvalho C.M.B., Cascardo J.C.M., Cavada B.S., Chueire L.M.O.,
 RA Creczynski-Pasa T.B., Cunha-Junior N.C., Fagundes N., Falcao C.L.,
 RA Fantinatti F., Farias I.P., Felipe M.S.S., Ferrari L.P., Ferro J.A.,
 RA Ferro M.I.T., Franco G.R., Freitas N.S.A., Furian L.R., T.B.,
 RA Gazzinelli R.T., Gomes E.A., Goncalves P.R., Grangeiro T.B.,
 RA Grattapaglia D., Grisard E.C., Hanna E.S., Jardim S.N., Laurino J.,
 RA Leoi L.C.T., Lima L.F.A., Loureiro M.F., Lyra M.C.C.P.,
 RA Madeira H.M.F., Manfio G.P., Maranhao A.O., Martins W.S.,
 RA di Mauro S.M.Z., de Medeiros S.R.B., Meisner R.V., Moreira M.A.M.,
 RA Nascimento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,
 RA Paixao R.F.C., Parente J.A., Pedrosa F.O., Pena S.D.J., Pereira J.O.,
 RA Pereira M., Pinto L.S.R.C., Pinto L.S., Porto J.I.R., Potrich D.P.,
 RA Ramalho-Neto C.E., Reis A.M.M., Rigo L.U., Rondinelli E.,
 RA Santos E.B.P., Santos F.R., Schneider M.P.C., Seunanez H.N.,
 RA Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,
 RA Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza K.R.L.,
 RA Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Urmenyi T.,
 RA Vettore A., Wassen R., Zaha A., Simpson A.J.G.;
 RT "The complete genome sequence of Chromobacterium violaceum reveals
 RT remarkable and exploitable bacterial adaptability";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).
 DR EMBL; AE016916; AAQ59371.1; --; Genomic_DNA.
 DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR008915; Peptidase M50.
 DR Pfam; PF02163; Peptidase_M50; 1.
 KW Complete proteome.
 SQ SEQUENCE 213 AA; 23777 MW; 1A941A4603F0AC1F CRC64;

Query Match 50.6%; Score 45; DB 2; Length 213;
 Best Local Similarity 75.0%; Pred. No. 32;
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 7 HINPTGVLLQL 18
 ||:| |||||
 Db 54 HIDPIGTVLLPL 65

Search completed: December 31, 2005, 23:47:04
 Job time : 8.24828 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model
Run on: December 31, 2005, 23:16:51 ; Search time 1.44828 Seconds
(without alignments)
1195.837 Million cell updates/sec

Title: US-10-602-220-29
Perfect score: 89
Sequence: 1 TSRLRSHINPTGTVLLQL 18
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89	100.0	429	1 GBHUA	alpha-galactosidase
2	52	58.4	809	2 G95299	probable adenylate
3	50	56.2	419	2 J45222	alpha-galactosidase
4	46	51.7	609	2 T35120	hypothetical prote
5	45	50.6	542	2 S45557	resb protein - bac
6	43	48.3	196	2 D81028	Mat/YceF/YhdE fami
7	43	48.3	286	2 J80138	gentamicin-N-acety
8	43	48.3	286	2 S06032	gentamycin 3'-N-ac
9	43	48.3	286	2 S06030	gentamycin 3'-N-ac
10	43	48.3	286	2 S09651	aminoglycoside N3'
11	43	48.3	286	2 S14541	aminoglycoside N3'
12	43	48.3	286	2 J20663	gentamicin resist
13	43	48.3	296	2 B71305	hypothetical prote
14	43	48.3	338	2 C75196	hypothetical prote
15	42	47.2	125	2 S43295	bone morphogenetic
16	42	47.2	151	2 S43296	bone morphogenetic
17	42	47.2	243	2 A33329	testis-specific pr
18	42	47.2	289	2 T34241	hypothetical prote
19	42	47.2	436	2 B55452	cartilage-derived
20	42	47.2	556	2 B83847	cytochrome c bioge
21	41	46.1	85	2 S04808	hypothetical prote
22	41	46.1	114	2 T10111	hypothetical prote
23	41	46.1	123	1 BKIP	biotin carboxyl ca
24	41	46.1	267	1 DCVKOP	biotidine-5'-phosp
25	41	46.1	347	2 A90318	hypothetical prote
26	41	46.1	361	2 S57895	hypothetical prote
27	41	46.1	495	2 S43294	bone morphogenetic
28	41	46.1	501	2 JC2347	growth/differentia
29	41	46.1	501	2 A55452	cartilage-derived

RESULT 1

GBHUA

alpha-galactosidase (EC 3.2.1.22) A precursor - human
N/Alternate names: alpha-D-galactoside galactohydrolase; melibiase
C/Species: Homo sapiens (man)
C/Date: 30-Jun-1987 #sequence_revision 27-Oct-1995 #text_change 09-Jul-2004
C/Accession: S04081; A29608; A30214; S14879; A00896; B00896; I37140
R/Kornreich, R.; Desnick, R.J.; Bishop, D.F.
Nucleic Acids Res. 17, 3301-3302, 1989.

A/Title: Nucleotide sequence of the human alpha-galactosidase A gene.
A/Reference number: S04081; MUID:89263745; PMID:2542896
A/Accession: S04081
A/Status: translation not shown

A/Molecule type: DNA

A/Residues: 1-429 <KOR>

A/Cross-references: UNIPROT:P06280; UNIPARC:UPI00000033A30; EMBL:X14448; NID:g11755; PIDN:Q14448

R/Quinn, M.; Hantzopoulos, P.; Fidanza, V.; Calhoun, D.H.
Gene 58, 177-188, 1987

A/Title: A genomic clone containing the promoter for the gene encoding the human lysosom

A/Reference number: A29608; MUID:88112869; PMID:2892762

A/Accession: A29608

A/Molecule type: DNA

A/Residues: 1-64 <QOI>

A/Cross-references: UNIPARC:UPI000016A96A; GB:M18242; NID:g182944; PIDN:AAA52514.1; PID:1

R/Bishop, D.F.; Kornreich, R.; Desnick, R.J.
Proc. Natl. Acad. Sci. U.S.A. 85, 3903-3907, 1988

A/Title: Structural organization of the human alpha-galactosidase A gene: further eviden

A/Reference number: A30214; MUID:88234528; PMID:2836863

A/Accession: A30214

A/Molecule type: DNA

A/Residues: 1-64 <BIS>

A/Cross-references: UNIPARC:UPI000016A96A; EMBL:M20317; EMBL:J03249

R/Koide, T.; Ishiura, M.; Iwai, K.; Inoue, M.; Kaneda, Y.; Okada, Y.; Uchida, T.
FEBS Lett. 259, 353-356, 1990

A/Title: A case of Fabry's disease in a patient with no alpha-galactosidase A activity c

A/Reference number: S14879; MUID:90092580; PMID:2152885

A/Accession: S14879

A/Molecule type: mRNA

A/Residues: 1, 'K', 3-39, 'S', 41-429 <KOI>

A/Cross-references: UNIPARC:UPI000011E285; EMBL:X16889

A/Experimental source: Fabry's disease patient

R/Bishop, D.F.; Calhoun, D.H.; Bernstein, H.S.; Hantzopoulos, P.; Quinn, M.; Desnick, R

Proc. Natl. Acad. Sci. U.S.A. 83, 4859-4863, 1986

A/Title: Human alpha-galactosidase A: nucleotide sequence of a cDNA clone encoding the r

A/Reference number: A00896; MUID:86259694; PMID:3014515

A/Accession: A00896

A/Molecule type: mRNA

A/Residues: 27-429 <B12>

A/Cross-references: UNIPARC:UPI0000000358; GB:M13571; NID:g178245; PIDN:AAA51676.1; PID

A/Experimental source: lung

A/Accession: B00896

A/Molecule type: protein

N;Alternate names: alpha-D-galactoside galactohydrolase
C;Species: Mus musculus (house mouse)
C;Date: 08-Feb-1996 #sequence_revision 19-Apr-1996 #text_change 09-Jul-2004
C;Accession: JC4522
R;Ohshima, T.; Murray, G.J.; Nagle, J.W.; Quirk, J.M.; Kraus, M.H.; Barton, N.W.; Brady, Gene 166, 277-280, 1995
A;Title: Structural organization and expression of the mouse gene encoding alpha-galactosidase
A;Reference number: JC4522; MUID:96125203; PMID:8543175
A;Accession: JC4522
A;Molecule type: mRNA
A;Residues: 1-419 <OHS>
A;Cross-references: UNIPROT:P51569; UNIPARC:UPI00000018CA; GB:U34071; NID:g1141787; PIDN: F/32-419/Product: alpha-galactosidase A #status predicted <MAT>
A;Experimental source: kidney, C57BL
C;Comment: This enzyme is a lysosomal enzyme that hydrolyses the alpha-D-galactosyl resi
C;Genetics:
A;Gene: alpha Gala
A;Introns: 65/2; 123/3; 183/1; 213/3; 267/3; 333/3
C;Superfamily: alpha-galactosidase
C;Keywords: glycoprotein; glycosidase; hydrolase; lysosome
F;1-31/Domain: signal sequence #status predicted <SIG>
F;32-419/Product: alpha-galactosidase A #status predicted <MAT>
F;139,192,215,408/Binding site: carboxydrate (Asn) #status predicted

Query Match	56.2%;	Score 50;	DB 2;	Length 419;
Best Local Similarity	50.0%;	Pred. No. 1.7;		
Matches	9;	Conservative	5;	Mismatches 4; Indels 0; Gaps 0;

QY 1 TSLRSHINPTGTVLQL 18
| : : : : :
DB 400 TLTLTRWPSGTVLFRL 417

RESULT 4
T25120
hypothetical protein T22C8.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T25120
R;Thomas, K.
submitted to the EMBL Data Library, April 1995
A;Reference number: Z19983
A;Accession: T25120
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-609 <WIL>
A;Cross-references: UNIPROT:Q22680; UNIPARC:UPI00000802C9; EMBL:Z49071; PIDN:CAA88879.1;
A;Experimental source: clone T22C8
C;Genetics:
A;Gene: CESP:T22C8.7
A;Map position: 2
A;Introns: 36/3; 66/2; 98/3; 146/2; 219/2; 274/3; 399/3; 475/1; 576/3

Query Match	51.7%;	Score 46;	DB 2;	Length 609;
Best Local Similarity	80.0%;	Pred. No. 12;		
Matches	8;	Conservative	1;	Mismatches 1; Indels 0; Gaps 0;

QY 4 LRSHINPTGT 13
:|:|||||

DB 43 IRLHINPTGT 52

```

RESULT 5
S4557
resB protein - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C:Accession: S45557; B69691
R:Sorokin, A.; Zunststein, E.; Azevedo, V.; Ehrlich, S.D.; Serror, P.
submitted to the EMBL Data Library, November 1993
A:Reference number: S45533
A:Accession: S45557
A>Status: preliminary
A:Molecule type: DNA

```


A;Residues: 1-542 <SOR>
 A;Cross-references: UNIPROT:P35161; UNIPARC:UPI0000060679; EMBL:L09228; NID:G410114; PID
 R;Kunst, F.; Ogasawara, N.; Mosser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
 C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.
 Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
 A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
 Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Ross, M.; Sadaie, Y.; Sato, T.; Scanlon,
 A;Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terptrata, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A;Reference number: A69580; MUID:98044033; PMID:9384377
 A;Accession: E69691
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-542 <KUN>
 A;Cross-references: UNIPARC:UPI0000060679; GB:Z99116; GB:AL009126; NID:G2634723; PIDN:CA
 A;Experimental source: strain 168
 C;Genetics:
 A;Gene: resB

Query Match 50.6%; Score 45; DB 2; Length 542;
 Best Local Similarity 72.7%; Pred. No. 16; Mismatches 2; Indels 0; Gaps 0;
 Matches 8; Conservative 1;

Qy 7 HINPTGTVLQ 17
 |||||
 Db 10 HINPVGTVLCE 20
 |||||

RESULT 6
 D81028
 Maf/YceF/YhdE family protein NMB1909 [imported] - *Neisseria meningitidis* (strain MC58 se
 C;Species: *Neisseria meningitidis*
 C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
 C;Accession: D81028
 R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
 xi, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizza, M.
 Science 287, 1809-1815, 2000
 A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
 A;Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
 A;Reference number: A81000; MUID:20175755; PMID:10710307
 A;Accession: D81028
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-196 <TE>
 A;Cross-references: UNIPROT:Q9JXS2; UNIPARC:UPI000013B498; GB:AE002540; GB:AE002098; NID
 A;Experimental source: serogroup B, strain MC58
 C;Genetics:
 A;Gene: NMB1909
 C;Superfamily: septum formation protein maf

Query Match 48.3%; Score 43; DB 2; Length 196;
 Best Local Similarity 50.0%; Pred. No. 11;
 Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 TSLRLSHINPTGTVLQL 18
 |||||
 Db 116 TGRMRRHIDKTVVNRQL 133
 |||||

RESULT 7
 JE0138
 gentamicin-N-acetyltransferase(3)-V - human
 C;Species: *Homo sapiens* (man)
 C;Date: 10-Jul-1998 #sequence_revision 10-Jul-1998 #text_change 26-May-2000

C;Accession: JE0138
 R;Fang, Y.; Song, H.Y.; Wang, F.; Zhu, D.Z.
 Chinese J. Microbiol. Immunol. 16, 409-412, 1996
 A;Title: Molecular cloning and characterization of gentamicin-N-acetyltransferase (3)-V
 A;Reference number: JE0138
 A;Accession: JE0138
 A;Molecule type: mRNA
 A;Residues: 1-286 <FAN>
 A;Cross-references: UNIPARC:UPI00001794CE
 C;Superfamily: *Escherichia coli* aminoglycoside N3'-acetyltransferase

Query Match 48.3%; Score 43; DB 2; Length 286;
 Best Local Similarity 66.7%; Pred. No. 17;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 LRSHINPTGTVL 15
 |||||
 Db 45 LRSVAGPTGTVM 56
 |||||

RESULT 8
 S06032

gentamycin 3'-N-acetyltransferase (EC 2.3.1.60) type IV - Enterobacteriaceae plasmid pW
 C;Species: plasmid pWP14
 C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 03-Jun-2002
 C;Accession: S06032
 R;Allmannsberger, R.; Braeu, B.; Piepersberg, W.
 Mol. Gen. Genet. 198, 514-520, 1985
 A;Title: Genes for gentamicin-(3)-N-acetyl-transferases III and IV. II. Nucleotide sequ
 A;Reference number: S06030; MUID:85239912; PMID:3892230
 A;Accession: S06032
 A;Molecule type: DNA
 A;Residues: 1-286 <ALL>
 A;Cross-references: UNIPARC:UPI0000000849; EMBL:X13542; NID:G45940; PIDN:CAA31893.1; PIC
 C;Genetics:
 A;Gene: aacC3
 A;Genome: plasmid pWP14
 C;Superfamily: *Escherichia coli* aminoglycoside N3'-acetyltransferase
 C;Keywords: acyltransferase; antibiotic resistance; coenzyme A

Query Match 48.3%; Score 43; DB 2; Length 286;
 Best Local Similarity 66.7%; Pred. No. 17;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 LRSHINPTGTVL 15
 |||||
 Db 45 LRSVAGPTGTVM 56
 |||||

RESULT 9
 S06030

gentamycin 3'-N-acetyltransferase (EC 2.3.1.60) type III - Enterobacteriaceae plasmid pW
 C;Species: Enterobacteriaceae
 C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 03-Jun-2002
 C;Accession: S06030
 R;Allmannsberger, R.; Braeu, B.; Piepersberg, W.
 Mol. Gen. Genet. 198, 514-520, 1985
 A;Title: Genes for gentamicin-(3)-N-acetyl-transferases III and IV. II. Nucleotide sequ
 A;Reference number: S06030; MUID:85239912; PMID:3892230
 A;Accession: S06030
 A;Molecule type: DNA
 A;Residues: 1-286 <ALL>
 A;Cross-references: UNIPARC:UPI000012508C; EMBL:X13543; NID:G45943; PIDN:CAA31895.1; PIC
 C;Genetics:
 A;Gene: aacC3
 A;Genome: plasmid pWP13a
 C;Superfamily: *Escherichia coli* aminoglycoside N3'-acetyltransferase
 C;Keywords: acyltransferase; antibiotic resistance; coenzyme A

Query Match 48.3%; Score 43; DB 2; Length 286;
 Best Local Similarity 66.7%; Pred. No. 17;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 LRSHINPTGTVL 15
| | | : | | | |
Db 45 LRSVAGPTGTVM 56

RESULT 10
S09651
aminoglycoside N3'-acetyltransferase (EC 2.3.1.81) isozyme II - Enterobacter cloacae
C:Species: Enterobacter cloacae
C>Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jul-2004
C:Accession: S09651
R:Vliegenthart, J.S.; Ketelaar-van Gaalen, P.A.G.; van de Klundert, J.A.M.
Antimicrob. Agents Chemother. 33, 1153-1159, 1989
A:Title: Nucleotide sequence of the aacC2 gene, a gentamicin resistance determinant invc
A:Reference number: S09651; MUID:90024972; PMID:2552900
A:Molecule type: DNA
A:Residues: 1-286 <VLI>
A:Cross-references: UNIPROT:P13245; UNIPARC:UPI0000000849; EMBL:X51534; NID:G40878; PIDN
C:Genetics:
A:Gene: aacC2
C:Superfamily: Escherichia coli aminoglycoside N3'-acetyltransferase
C:Keywords: acyltransferase; coenzyme A

Query Match 48.3%; Score 43; DB 2; Length 286;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 LRSHINPTGTVL 15
| | | : | | | |
Db 45 LRSVAGPTGTVM 56

RESULT 11
S14541
aminoglycoside N3'-acetyltransferase (EC 2.3.1.81) - Escherichia coli
C:Species: Escherichia coli
C>Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C:Accession: S14541
R:Yakulenko, S.B.; Entina, H.
submitted to the EMBL Data Library, October 1990
A:Description: Nucleotide sequence of the aacC2 gene from E. coli R-plasmid.
A:Reference number: S14541
A:Accession: S14541
A:Molecule type: DNA
A:Residues: 1-286 <VAK>
A:Cross-references: UNIPROT:Q03634; UNIPARC:UPI000000B81C6; EMBL:X54723; NID:G45769; PIDN
C:Superfamily: Escherichia coli aminoglycoside N3'-acetyltransferase
C:Keywords: acyltransferase; coenzyme A

Query Match 48.3%; Score 43; DB 2; Length 286;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 LRSHINPTGTVL 15
| | | : | | | |
Db 45 LRSVAGPTGTVM 56

RESULT 12
JC2063
gentamicin resistance determinant - Serratia marcescens
C:Species: Serratia marcescens
C>Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 09-Jul-2004
C:Accession: JC2063
R:Jung, J.S.; Cheong, T.C.; Cho, M.S.; Hah, Y.C.; Chung, J.H.
Biochem. Biophys. Res. Commun. 198, 1084-1089, 1994
A:Title: Nucleotide sequence and expression of a gentamicin resistance gene isolated fro
A:Reference number: JC2063; MUID:94161718; PMID:8117265
A:Accession: JC2063
A:Molecule type: DNA
A:Residues: 1-286 <JUN>
A:Cross-references: UNIPROT:Q52310; UNIPARC:UPI000000B39F5

C:Superfamily: Escherichia coli aminoglycoside N3'-acetyltransferase

Query Match 48.3%; Score 43; DB 2; Length 286;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 LRSHINPTGTVL 15
| | | : | | | |
Db 45 LRSVAGPTGTVM 56

RESULT 13
B71305
hypothetical protein TP0608 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 09-Jul-2004
C:Accession: B71305
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uterback, T.; McDo
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770; PMID:9665876
A:Accession: B71305
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-296 <COL>
A:Cross-references: UNIPROT:O83617; UNIPARC:UPI0000139B22; GB:AE001235; GB:AE000520; NID
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0608
C:Superfamily: syphilis spirochete hypothetical protein TP0608

Query Match 48.3%; Score 43; DB 2; Length 296;
Best Local Similarity 52.9%; Pred. No. 17;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 SRLRSHINPTGTVLLQL 18
| | | | | | | | | | | | | | | | | |
Db 230 SRKFSAINPTGSLALEI 246

RESULT 14
C75196
hypothetical protein PAB0054 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C:Accession: C75196
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: C75196
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-338 <KAW>
A:Cross-references: UNIPROT:Q9V2I0; UNIPARC:UPI00000034415; GB:AJ248283; GB:AL096836; NID
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB0054
C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0054

Query Match 48.3%; Score 43; DB 2; Length 338;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 RLRSINPTGTVLLQL 18
: | | | | | | | | | | | | | | | | | |
Db 87 QIRSLNPKMTMLTYL 102

RESULT 15
S43295

bone morphogenetic protein homolog GDF6 precursor - mouse (fragment)
 N:Alternate names: growth and differentiation factor 6
 C:Species: Mus musculus (house mouse)
 C:Date: 20-Oct-1994 #sequence_revision 07-Feb-1997 #text_change 09-Jul-2004
 C:Accession: S43295
 R:Storm, E.E.; Huynh, T.V.; Copeland, N.G.; Jenkins, N.A.; Kingsley, D.M.; Lee, S.J.
 Nature 368, 639-643, 1994
 A:Title: Limb alterations in brachypodism mice due to mutations in a new member of the TGF-beta family
 A:Reference number: S43294; MUID:94195427; PMID:8145850
 A:Accession: S43295
 A:Molecule type: DNA
 A:Residues: 1-125 <STO>
 A:Cross-references: UNIPROT:P43028; UNIPARC:UPI0000024471; EMBL:U08338; NID:9488463; PIR:G488463
 C:Genetics:
 C:Superfamily: inhibin
 F:1-5/Domain: polybasic protease recognition site #status predicted <PPR>
 F:6-125/Product: bone morphogenetic protein homolog GDF6 (fragment) #status predicted <M>

Query Match	47.2%;	Score 42;	DB 2;	Length 125;
Best Local Similarity	50.0%;	Pred. No. 10;		
Matches	7;	Conservative	3;	Mismatches 4;
				Indels 0;
				Gaps 0;

QY 4 LRSHINPTGVLLQ 17
 ||||: || :||
 Db 61 LRSHLEPTNHAIQ 74

Search completed: December 31, 2005, 23:47:44
 Job time : 3.44828 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 21:05:01 ; Search time 6.57931 Seconds
(without alignments)
1202.074 Million cell updates/sec

Title: US-10-602-220-29

Perfect score: 89

Sequence: 1 TSRLRSHINPTGTVLLQL 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_21.*

- 1: geneseqp1980s.*
- 2: geneseqp1990s.*
- 3: geneseqp2000s.*
- 4: geneseqp2001s.*
- 5: geneseqp2002s.*
- 6: geneseqp2003as.*
- 7: geneseqp2003bs.*
- 8: geneseqp2004s.*
- 9: geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89	100.0	18	5	Aae19258 Human rec
2	89	100.0	18	7	Add84768 Alpha-gal
3	89	100.0	18	8	Adj88294 Tobacco m
4	89	100.0	18	8	Adm48698 Human rga
5	89	100.0	18	8	Adu66933 Human alp
6	89	100.0	18	9	Aea27462 Human rga
7	89	100.0	22	5	Aae19256 Human rec
8	89	100.0	22	8	Add84766 Alpha-gal
9	89	100.0	22	8	Adj88292 Tobacco m
10	89	100.0	22	8	Adm48696 Human rga
11	89	100.0	22	8	Adu66931 Human alp
12	89	100.0	22	9	Aea27460 Human rga
13	89	100.0	24	5	Aae19259 Human rec
14	89	100.0	24	7	Add84769 Alpha-gal
15	89	100.0	24	8	Adj88295 Tobacco m
16	89	100.0	24	8	Adm48699 Human rga
17	89	100.0	24	8	Adu66934 Human alp
18	89	100.0	24	9	Aea27463 Human rga
19	89	100.0	25	9	Aea27477 Human rga
20	89	100.0	26	5	Aae19254 Human rec
21	89	100.0	26	7	Add84764 Alpha-gal
22	89	100.0	26	8	Adj88290 Tobacco m
23	89	100.0	26	8	Adm48694 Human rga
24	89	100.0	26	8	Adu66929 Human alp

RESULT 1

AAE19258	89	100.0	26	9	AEA27458	Human rga
ID	AAE19258	standard; peptide; 18 AA.	28	5	AAE19257	Human rec
XX	AAE19258;		28	9	AEA27479	Human rga
AC	AAE19258;		30	5	AAE19252	Human rec
XX	21-MAY-2002 (first entry)		30	7	ADD84762	Alpha-gal
DT	Human recombinant alpha-galactosidase A modified peptide, rGal-12.		30	8	ADJ88288	Tobacco m
DE	Glucocerebrosidase; alpha-galactosidase; Tay-Sachs disease; plant; enzyme replacement therapy; Niemann-Pick disease; Gaucher disease; Fabry disease; lysosomal storage disease; human.		30	8	ADM48692	Human rga
XX	Homo sapiens.		30	8	ADU66927	Human alp
OS	Synthetic.		30	8	ADU66927	Human alp
OS			30	8	ADU66927	Human alp
PN	WO200208404-A2.		30	8	ADU66927	Human alp
XX	31-JAN-2002.		30	8	ADU66927	Human alp
PD			30	8	ADU66927	Human alp
XX	26-JUL-2001; 2001WO-US024111.		30	8	ADU66927	Human alp
PF	26-JUL-2000; 2000US-00626127.		30	8	ADU66927	Human alp
XX	(LARG-) LARGE SCALE BIOLOGY CORP.		30	8	ADU66927	Human alp
PR	Garger SJ, Turpen TH, Kumagai MH;		30	8	ADU66927	Human alp
XX	WPI; 2002-195873/25.		30	8	ADU66927	Human alp
XX	New glucocerebrosidase and alpha-galactosidase having a post-translational modification, useful in enzyme replacement therapy for treating lysosomal storage diseases, e.g. Gaucher disease, Niemann-Pick disease, Fabry disease.		30	8	ADU66927	Human alp
XX	Example 11; Fig 5; 102pp; English.		30	8	ADU66927	Human alp
XX	The invention relates to production of lysosomal enzymes in plants by transient expression. The invention particularly relates to glucocerebrosidase and alpha-galactosidase having a post-translational modification. The enzymes are useful in enzyme replacement therapy for treating lysosomal storage diseases (e.g. Gaucher disease, Niemann-Pick disease, Fabry disease and Tay-Sachs disease), in researches for developing new approaches to medical treatment of lysosomal storage diseases and in industrial processes involving enzymatic substrate hydrolysis. The present sequence is human recombinant alpha-		30	8	ADU66927	Human alp

```
CC galactosidase-A C-terminal modified peptide
XX
SQ Sequence 18 AA;

Query Match          100.0%; Score 89; DB 5; Length 18;
Best Local Similarity 100.0%; Pred. No. 5.4e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGVLLQL 18
   |||||
Db 1 TSRLRSHINPTGVLLQL 18
   |||||

RESULT 2
ADD84768
ID ADD84768 standard; peptide; 18 AA.
XX
AC ADD84768;
XX
DT 29-JAN-2004 (first entry)
XX
DE Alpha-galactosidase derivative peptide fragment #7.
XX
KW Tobacco mosaic virus; alpha-galactosidase; lysosomal enzyme;
KW enzyme replacement therapy; lysosomal disease; enzyme.
XX
OS Tobacco mosaic virus.
XX
PN US2003106095-A1.
PD
PP 05-JUN-2003.
XX
PF 20-MAR-2002; 2002US-00103327.
XX
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX
PA (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
XX
PI Garger SJ, Turpen TH, Kumagai MH;
XX
DR WPI; 2003-801257/75.
XX
XX
XX New polynucleotide for producing active recombinant human and animal
XX lysosomal enzymes in a plant expression system that can be used in enzyme
XX replacement therapy.
XX
PS Example 11; SEQ ID NO 29; 77pp; English.
XX
CC The invention relates to human alpha-galactosidase derivatives and the
CC nucleic acids encoding them. The polypeptides are used in a method for
CC producing active recombinant human and animal lysosomal enzymes in a
CC plant expression system. The enzymes can be used in enzyme replacement
CC therapy for the therapeutic treatment of human and animal lysosomal
CC diseases. This sequence represents an alpha-galactosidase derivative
CC peptide used in the method of the invention.
XX
SQ Sequence 18 AA;

Query Match          100.0%; Score 89; DB 7; Length 18;
Best Local Similarity 100.0%; Pred. No. 5.4e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGVLLQL 18
   |||||
Db 1 TSRLRSHINPTGVLLQL 18
   |||||

RESULT 3
ADJ88294
ID ADJ88294 standard; peptide; 18 AA.
XX
AC ADJ88294;
XX
DT 06-MAY-2004 (first entry)
XX
DE Tobacco mosaic virus WT rGAL-12 (galactosidase) C-terminal peptide.
XX
KW Galactosidase; GAL; gene therapy; lysosomal storage disease;
KW Fabry's disease; Gaucher's disease.
XX
OS Tobacco mosaic virus.
XX
PN US2004016021-A1.
PD
PP 22-JAN-2004.
XX
PF 23-JUN-2003; 2003US-00602219.
XX
PR 26-FEB-1988; 88US-00160766.
PR 26-FEB-1988; 88US-00160771.
PR 15-JUL-1988; 88US-00219279.
PR 17-FEB-1989; 89US-00310861.
PR 05-MAY-1989; 89US-00347637.
PR 08-JUN-1989; 89US-00363138.
PR 22-OCT-1990; 90US-00600244.
PR 16-JAN-1991; 91US-00641617.
PR 26-JUL-1991; 91US-00737899.
PR 01-AUG-1991; 91US-00739143.
PR 31-JUL-1992; 92US-00923692.
PR 30-DEC-1992; 92US-00997733.
PR 29-DEC-1993; 93US-00176414.
PR 19-JAN-1994; 94US-00184237.
PR 14-OCT-1994; 94US-00324003.
PR 21-MAY-1999; 99US-00316572.
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX
PA (TURP/) TURPEN T H.
PA (FOGU/) FOGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX
PI Turpen TH, Fogue GP, Erwin RL, Grill LK;
XX
DR WPI; 2004-108227/11.
XX
XX
XX New lysosomal enzymes, useful in treating human and animal lysosomal
XX storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
PS Example 11; SEQ ID NO 29; 71pp; English.
XX
CC The invention relates to nucleotide encoding galactosidase (GAL). The
CC invention is useful in gene therapy. The polynucleotides and polypeptides
CC are useful in treating human and animal lysosomal storage diseases, e.g.
CC Fabry's disease and Gaucher's diseases. The present sequence is Tobacco
CC mosaic virus galactosidase C-terminal peptide.
XX
SQ Sequence 18 AA;

Query Match          100.0%; Score 89; DB 8; Length 18;
Best Local Similarity 100.0%; Pred. No. 5.4e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGVLLQL 18
   |||||
Db 1 TSRLRSHINPTGVLLQL 18
   |||||

RESULT 4
ADM48698
ID ADM48698 standard; peptide; 18 AA.
XX
AC ADM48698;
```

```
XX 03-JUN-2004 (first entry)
XX Human rGAL-12 C-terminal peptide.
XX
XX Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
XX lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
XX Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
XX enzyme.
XX
XX Homo sapiens.
XX
XX US2004023281-A1.
XX
XX 05-FEB-2004.
XX
XX 23-JUN-2003; 2003US-00602220.
XX
XX 26-FEB-1988; 88US-00160766.
XX 26-FEB-1988; 88US-00160771.
XX 17-FEB-1989; 89US-00310881.
XX 22-OCT-1990; 90US-00600244.
XX 31-JUL-1992; 92US-00923692.
XX 30-DEC-1992; 92US-00997733.
XX 29-DEC-1993; 93US-00176414.
XX 19-JAN-1994; 94US-00184237.
XX 14-OCT-1994; 94US-00324003.
XX 21-MAY-1999; 99US-00316572.
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.
XX
XX (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX (POGU/) POGUE G P.
XX (ERWI/) ERWIN R L.
XX (GRIL/) GRILL L K.
XX
XX Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
XX WPI; 2004-142650/14.
XX
XX New alpha-galactosidase polypeptides, useful in producing recombinant
XX lysosomal enzymes for the treatment of lysosomal storage diseases, such
XX as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
XX disease.
XX
XX Claim 1; SEQ ID NO 29; 72bp; English.
XX
XX The present invention relates to novel galactosidase (Gal) proteins such
XX as rGAL-12, rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions
XX of the present invention are useful for producing recombinant lysosomal
XX enzymes for enzyme replacement therapy for treating human and animal
XX lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
XX disease, Fabry disease and Tay-Sachs disease. The present sequence is
XX human rGAL-12 C-terminal peptide used in the exemplification of the
XX invention.
XX
XX Sequence 18 AA;
XX
XX Query Match 100.0%; Score 89; DB 8; Length 18;
XX Best Local Similarity 100.0%; Pred. No. 5.4e-08;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 TSRLRSHINPTGVLLQL 18
XX |||||
XX 1 TSRLRSHINPTGVLLQL 18
XX
XX RESULT 5
XX ADU66933
XX ID ADU66933 standard; peptide; 18 AA.
XX
XX ADU66933;
XX
XX Query Match 100.0%; Score 89; DB 8; Length 18;
XX Best Local Similarity 100.0%; Pred. No. 5.4e-08;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 TSRLRSHINPTGVLLQL 18
XX |||||
XX 1 TSRLRSHINPTGVLLQL 18
XX
XX RESULT 6
XX AEA27462
XX ID AEA27462 standard; peptide; 18 AA.
XX
XX AEA27462;
XX
XX 11-AUG-2005 (first entry)
XX
XX Human rGAL-12 C-terminal peptide, SEQ ID NO: 29.
XX
XX Gauchers disease; metabolic; neurological disease; niemann pick disease;
XX genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
XX antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase.
XX
XX Homo sapiens.
XX
```

10-FEB-2005 (first entry)

Human alpha-galactosidase-A C-terminal peptide #7.

Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase; enzyme replacement therapy; lysosomal storage disease; Gaucher's disease; Niemann-Pick disease; Fabry's disease; Tay-Sachs disease; Hurler's syndrome; Hurler-Scheie syndrome; nephrotrophic; human.

Homo sapiens.

US2004234516-A1.

25-NOV-2004.

21-MAY-2004; 2004US-00851388.

26-JUL-2000; 2000US-00626127.

13-NOV-2001; 2001US-00993059.

20-MAR-2002; 2002US-00103327.

(LARG-) LARGE SCALE BIOLOGY CORP.

Garger SJ, Turpen TH, Kumagai MH; WPI; 2004-821274/81.

A pharmaceutical composition comprising a lysosomal enzyme, useful for enzyme replacement therapy for the treatment of lysosomal storage diseases, such as Fabry's disease.

Example 11; SEQ ID NO 29; 88bp; English.

The present invention relates to the production of human and animal lysosomal enzymes in plants by a transient plant expression system. The invention relates to glucocerebrosidase (GCB, GCR) and alpha-galactosidase enzymes having a post-translational modification provided by the plant expression system. The invention is useful in enzyme replacement therapy for treating lysosomal storage diseases such as Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is also useful in researches for developing new approaches to medical treatment of lysosomal storage diseases and in industrial processes involving enzymatic substrate hydrolysis. The present sequence is the human alpha-galactosidase-A C-terminal peptide.

Sequence 18 AA;

Query Match 100.0%; Score 89; DB 8; Length 18; Best Local Similarity 100.0%; Pred. No. 5.4e-08; Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TSRLRSHINPTGVLLQL 18
|||||

1 TSRLRSHINPTGVLLQL 18

RESULT 6

AEA27462

ID AEA27462 standard; peptide; 18 AA.

AEA27462;

11-AUG-2005 (first entry)

Human rGAL-12 C-terminal peptide, SEQ ID NO: 29.

Gauchers disease; metabolic; neurological disease; niemann pick disease; genetic disorder; Fabry disease; metabolic disorder; tay sachs disease; antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase.

Homo sapiens.

```

XX PN US2005125859-A1.
XX PD 09-JUN-2005.
XX PF 08-NOV-2004; 2004US-00984389.
XX PR 26-JUL-2000; 2000US-00626127.
XX PR 13-NOV-2001; 2001US-00993059.
XX PR 20-MAR-2002; 2002US-00103327.
XX PA (LARG-) LARGE SCALE BIOLOGY CORP.
XX PI Garger SJ, Turpen TH, Kumagai MH;
XX DR WPI; 2005-404004/41.
XX PT New isolated polypeptides useful for producing lysosomal enzymes in
XX PT plants to be utilized in enzyme replacement therapy or for the
XX PT therapeutic treatment of human or animal lysosomal storage diseases, e.g.
XX PT Gaucher's disease.
XX PS Example 11; SEQ ID NO 29; 88pp; English.
XX CC The present invention relates to the production of human and animal
XX CC lysosomal enzymes in plants by a transient plant expression system. The
XX CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
XX CC galactosidase (Gal) enzymes having a post-translational modification
XX CC provided by the plant expression system. The invention is useful in
XX CC enzyme replacement therapy for treating lysosomal storage diseases such
XX CC as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
XX CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
XX CC also useful in researches for developing new approaches to medical
XX CC treatment of lysosomal storage diseases and in industrial processes
XX CC involving enzymatic substrate hydrolysis. The present sequence is the
XX CC human galactosidase (GAL) C-terminal peptide present in the vector rGAL-
XX CC 12.
XX SQ Sequence 18 AA;
XX Query Match 100.0%; Score 89; DB 9; Length 18;
XX Best Local Similarity 100.0%; Pred. No. 5.4e-08;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18
RESULT 7
AAE19256
ID AAE19256 standard; peptide; 22 AA.
XX AC AAE19256;
XX PD 21-MAY-2002 (first entry)
XX DT Human recombinant alpha-galactosidase A modified peptide, rGal-8.
XX DE Glucocerebrosidase; alpha-galactosidase; Tay-Sachs disease; plant;
XX KW enzyme replacement therapy; Niemann-Pick disease; Gaucher disease;
XX KW Fabry disease; lysosomal storage disease; human.
XX OS Homo sapiens.
XX OS Synthetic.
XX PN WO200208404-A2.
XX PD 31-JAN-2002.
XX PR 26-JUL-2001; 2001WO-US024111.
XX PR 26-JUL-2000; 2000US-00626127.

```

```

XX PA (LARG-) LARGE SCALE BIOLOGY CORP.
XX PI Garger SJ, Turpen TH, Kumagai MH;
XX DR WPI; 2002-195873/25.
XX PR New glucocerebrosidase and alpha-galactosidase having a post-
XX PR translational modification, useful in enzyme replacement therapy for
XX PR treating lysosomal storage diseases, e.g. Gaucher disease, Niemann-Pick
XX PR disease, Fabry disease.
XX PS Example 11; Fig 5; 102pp; English.
XX CC The invention relates to production of lysosomal enzymes in plants by
XX CC transient expression. The invention particularly relates to
XX CC glucocerebrosidase and alpha-galactosidase having a post-translational
XX CC modification. The enzymes are useful in enzyme replacement therapy for
XX CC treating lysosomal storage diseases (e.g. Gaucher disease, Niemann-Pick
XX CC disease, Fabry disease and Tay-Sachs disease), in researches for
XX CC developing new approaches to medical treatment of lysosomal storage
XX CC diseases and in industrial processes involving enzymatic substrate
XX CC hydrolysis. The present sequence is human recombinant alpha-
XX CC galactosidase-A C-terminal modified peptide
XX SQ Sequence 22 AA;
XX Query Match 100.0%; Score 89; DB 5; Length 22;
XX Best Local Similarity 100.0%; Pred. No. 6.8e-08;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18
RESULT 8
ADD84766
ID ADD84766 standard; peptide; 22 AA.
XX AC ADD84766;
XX DT 29-JAN-2004 (first entry)
XX DE Alpha-galactosidase derivative peptide fragment #5.
XX KW Tobacco mosaic virus; alpha-galactosidase; lysosomal enzyme;
XX KW enzyme replacement therapy; lysosomal disease; enzyme.
XX OS Tobacco mosaic virus.
XX PN US2003106095-A1.
XX PD 05-JUN-2003.
XX PF 20-MAR-2002; 2002US-00103327.
XX PR 26-JUL-2000; 2000US-00626127.
XX PR 13-NOV-2001; 2001US-00993059.
XX PA (GARG/) GARGER S J.
XX PA (TURP/) TURPEN T H.
XX PA (KUMA/) KUMAGAI M H.
XX PI Garger SJ, Turpen TH, Kumagai MH;
XX DR WPI; 2003-801257/75.
XX PT New polynucleotide for producing active recombinant human and animal
XX PT lysosomal enzymes in a plant expression system that can be used in enzyme
XX PT replacement therapy.
XX PS Example 11; SEQ ID NO 27; 77pp; English.

```


XX CC The invention relates to human alpha-galactosidase derivatives and the
CC nucleic acids encoding them. The polypeptides are used in a method for
CC producing active recombinant human and animal lysosomal enzymes in a
CC plant expression system. The enzymes can be used in enzyme replacement
CC therapy for the therapeutic treatment of human and animal lysosomal
CC diseases. This sequence represents an alpha-galactosidase derivative
CC peptide used in the method of the invention.

XX SQ Sequence 22 AA;

Query Match 100.0%; Score 89; DB 7; Length 22;
Best Local Similarity 100.0%; Pred. No. 6.8e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGTVLLQL 18
|||||
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 9
ADJ88292
ID ADJ88292 standard; peptide; 22 AA.
XX AC ADJ88292;
XX DT 06-MAY-2004 (first entry)
XX DE Tobacco mosaic virus WT rGAL-8 (galactosidase) C-terminal peptide.
XX KW Galactosidase; GAL; gene therapy; lysosomal storage disease;
KW Fabry's disease; Gaucher's disease.
XX OS Tobacco mosaic virus.
XX PN US2004016021-A1.
XX PD 22-JAN-2004.
XX PF 23-JUN-2003; 2003US-00602219.
XX PR 26-FEB-1988; 88US-00160766.
PR 26-FEB-1988; 88US-00160771.
PR 15-JUL-1988; 88US-00219279.
PR 17-FEB-1989; 89US-00310881.
PR 05-MAY-1989; 89US-00347637.
PR 08-JUN-1989; 89US-00363138.
PR 22-OCT-1990; 90US-00600244.
PR 16-JAN-1991; 91US-00641617.
PR 26-JUL-1991; 91US-00737899.
PR 01-AUG-1991; 91US-00739143.
PR 31-JUL-1992; 92US-00923692.
PR 30-DEC-1992; 92US-00997733.
PR 29-DEC-1993; 93US-00176414.
PR 19-JAN-1994; 94US-00184237.
PR 14-OCT-1994; 94US-00324003.
PR 21-MAY-1999; 99US-00316572.
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX (TURP/) TURPEN T H.
PA (POGU/) POGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX Turpen TH, Pogue GP, Erwin RL, Grill LK;
XX WPI; 2004-108227/11.
XX New lysosomal enzymes, useful in treating human and animal lysosomal
PT storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX Example 11; SEQ ID NO 27; 71pp; English.

XX CC The invention relates to nucleotide encoding galactosidase (GAL). The
CC invention is useful in gene therapy. The polynucleotides and polypeptides
CC are useful in treating human and animal lysosomal storage diseases, e.g.
CC Fabry's disease and Gaucher's disease. The present sequence is Tobacco
CC mosaic virus galactosidase C-terminal peptide.

XX SQ Sequence 22 AA;

Query Match 100.0%; Score 89; DB 8; Length 22;
Best Local Similarity 100.0%; Pred. No. 6.8e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGTVLLQL 18
|||||
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 10
ADM48696
ID ADM48696 standard; peptide; 22 AA.
XX AC ADM48696;
XX DT 03-JUN-2004 (first entry)
XX DE Human rGAL-8 C-terminal peptide.
XX KW Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
KW enzyme.
XX OS Homo sapiens.
XX PN US2004023281-A1.
XX PD 05-FEB-2004.
XX PF 23-JUN-2003; 2003US-00602220.
XX PR 26-FEB-1988; 88US-00160766.
PR 26-FEB-1988; 88US-00160771.
PR 17-FEB-1989; 89US-00310881.
PR 22-OCT-1990; 90US-00600244.
PR 31-JUL-1992; 92US-00923692.
PR 30-DEC-1992; 92US-00997733.
PR 29-DEC-1993; 93US-00176414.
PR 19-JAN-1994; 94US-00184237.
PR 14-OCT-1994; 94US-00324003.
PR 21-MAY-1999; 99US-00316572.
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
PA (POGU/) POGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
XX WPI; 2004-142650/14.
XX New alpha-galactosidase polypeptides, useful in producing recombinant
PT lysosomal enzymes for the treatment of lysosomal storage diseases, such
as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
disease.
XX Example 11; SEQ ID NO 27; 72pp; English.
XX The present invention relates to novel galactosidase (Gal) proteins such
as rGAL-12, rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions

CC of the present invention are useful for producing recombinant lysosomal
CC enzymes for enzyme replacement therapy for treating human and animal
CC lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
CC disease, Fabry disease and Tay-Sachs disease. The present sequence is
CC human rGAL-8 C-terminal peptide used in the exemplification of the
CC invention.

XX
XX
SQ Sequence 22 AA;

Query Match 100.0%; Score 89; DB 8; Length 22;
Best Local Similarity 100.0%; Pred. No. 6.8e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGVLLQL 18
| | | | | | | | | | | | | | | | | |
DB 1 TSRLRSHINPTGVLLQL 18

RESULT 11
ADU66931
ID ADU66931 standard; peptide; 22 AA.

AC ADU66931;
XX
XX 10-FEB-2005 (first entry)
XX
XX Human alpha-galactosidase-A C-terminal peptide #5.

XX Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase;
KW enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;
KW Niemann-Pick disease; Fabry's disease; Tay-Sachs disease;
KW Hurler's syndrome; Hurler-Scheie syndrome; nephrotropic; human.

XX Homo sapiens.

XX US2004234516-A1.

XX 25-NOV-2004.

XX 21-MAY-2004; 2004US-00851389.

XX 26-JUL-2000; 2000US-00626127.

PR 13-NOV-2001; 2001US-00993059.

PR 20-MAR-2002; 2002US-00103327.

XX (LARG-) LARGE SCALE BIOLOGY CORP.

XX Garger SJ, Turpen TH, Kumagai MH;

XX WPI; 2004-821274/81.

XX A pharmaceutical composition comprising a lysosomal enzyme, useful for
PT enzyme replacement therapy for the treatment of lysosomal storage
PT diseases, such as Fabry's disease.

XX Example 11; SEQ ID NO 27; 88pp; English.

XX The present invention relates to the production of human and animal
CC lysosomal enzymes in plants by a transient plant expression system. The
CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
CC galactosidase enzymes having a post-translational modification provided
CC by the plant expression system. The invention is useful in enzyme
CC replacement therapy for treating lysosomal storage diseases such as
CC Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs
CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
CC also useful in researches for developing new approaches to medical
CC treatment of lysosomal storage diseases and in industrial processes
CC involving enzymatic substrate hydrolysis. The present sequence is the
CC human alpha-galactosidase-A C-terminal peptide.

XX Sequence 22 AA;

Query Match 100.0%; Score 89; DB 8; Length 22;

Best Local Similarity 100.0%; Pred. No. 6.8e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGVLLQL 18
| | | | | | | | | | | | | | | | | |
DB 1 TSRLRSHINPTGVLLQL 18

RESULT 12
AEA27460
ID AEA27460 standard; peptide; 22 AA.

XX AEA27460;

XX 11-AUG-2005 (first entry)

XX Human rGAL-8 C-terminal peptide, SEQ ID NO: 27.

XX Gauchers disease; metabolic; neurological disease; niemann pick disease;
KW genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
KW antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase.

XX Homo sapiens.

XX US2005125859-A1.

XX 09-JUN-2005.

XX 08-NOV-2004; 2004US-00984389.

XX 26-JUL-2000; 2000US-00626127.

PR 13-NOV-2001; 2001US-00993059.

PR 20-MAR-2002; 2002US-00103327.

XX (LARG-) LARGE SCALE BIOLOGY CORP.

XX Garger SJ, Turpen TH, Kumagai MH;

XX WPI; 2005-404004/41.

XX New isolated polypeptides useful for producing lysosomal enzymes in
PT plants to be utilized in enzyme replacement therapy or for the
PT therapeutic treatment of human or animal lysosomal storage diseases, e.g.
PT Gaucher's disease.

XX Claim 22; SEQ ID NO 27; 88pp; English.

XX The present invention relates to the production of human and animal
CC lysosomal enzymes in plants by a transient plant expression system. The
CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
CC galactosidase (Gal) enzymes having a post-translational modification
CC provided by the plant expression system. The invention is useful in
CC enzyme replacement therapy for treating lysosomal storage diseases such
CC as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
CC also useful in researches for developing new approaches to medical
CC treatment of lysosomal storage diseases and in industrial processes
CC involving enzymatic substrate hydrolysis. The present sequence is the
CC human galactosidase (GAL) C-terminal peptide present in the vector rGAL-
CC 8.

XX Sequence 22 AA;

Query Match 100.0%; Score 89; DB 9; Length 22;

Best Local Similarity 100.0%; Pred. No. 6.8e-08; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGVLLQL 18
| | | | | | | | | | | | | | | | | |
DB 1 TSRLRSHINPTGVLLQL 18

RESULT 13

```
AAE19259
ID AAE19259 standard; peptide; 24 AA.
XX
AC AAE19259;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human recombinant alpha-galactosidase A modified peptide, rGal-12R.
XX
KW Glucocerebrosidase; alpha-galactosidase; Tay-Sachs disease; plant;
KW enzyme replacement therapy; Niemann-Pick disease; Gaucher disease;
KW Fabry disease; lysosomal storage disease; human.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200208404-A2.
XX
PD 31-JAN-2002.
XX
PF 26-JUL-2001; 2001WO-US024111.
XX
PR 26-JUL-2000; 2000US-00626127.
XX
PA (LARG-) LARGE SCALE BIOLOGY CORP.
XX
PI Garger SJ, Turpen TH, Kumagai MH;
XX
DR WPI; 2002-195873/25.
XX
PT New glucocerebrosidase and alpha-galactosidase having a post-
PT translational modification, useful in enzyme replacement therapy for
PT treating lysosomal storage diseases, e.g. Gaucher disease, Niemann-Pick
PT disease, Fabry disease.
XX
PS Example 11; Fig 5; 102pp; English.
XX
CC The invention relates to production of lysosomal enzymes in plants by
CC transient expression. The invention particularly relates to
CC glucocerebrosidase and alpha-galactosidase having a post-translational
CC modification. The enzymes are useful in enzyme replacement therapy for
CC treating lysosomal storage diseases (e.g. Gaucher disease, Niemann-Pick
CC disease, Fabry disease and Tay-Sachs disease), in researches for
CC developing new approaches to medical treatment of lysosomal storage
CC diseases and in industrial processes involving enzymatic substrate
CC hydrolysis. The present sequence is human recombinant alpha-
CC galactosidase-A C-terminal modified peptide
XX
SQ Sequence 24 AA;

Query Match 100.0%; Score 89; DB 5; Length 24;
Best Local Similarity 100.0%; Pred. No. 7.5e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGVLLQL 18
Db 1 TSRLRSHINPTGVLLQL 18

RESULT 14
ADD84769
ID ADD84769 standard; peptide; 24 AA.
XX
AC ADD84769;
XX
DT 29-JAN-2004 (first entry)
XX
DE Alpha-galactosidase derivative peptide fragment #8.
XX
KW Tobacco mosaic virus; alpha-galactosidase; lysosomal enzyme;
KW enzyme replacement therapy; lysosomal disease; enzyme.
XX
OS Tobacco mosaic virus.

Query Match 100.0%; Score 89; DB 5; Length 24;
Best Local Similarity 100.0%; Pred. No. 7.5e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGVLLQL 18
Db 1 TSRLRSHINPTGVLLQL 18

RESULT 15
ADJ88295
ID ADJ88295 standard; peptide; 24 AA.
XX
AC ADJ88295;
XX
DT 06-MAY-2004 (first entry)
XX
DE Tobacco mosaic virus WT rGal-12R (galactosidase) C-terminal peptide.
XX
KW Galactosidase; GAL; gene therapy; lysosomal storage disease;
KW Fabry's disease; Gaucher's disease.
XX
OS Tobacco mosaic virus.
XX
PN US2004016021-A1.
XX
PD 22-JAN-2004.
XX
PF 23-JUN-2003; 2003US-00602219.
XX
PR 26-FEB-1988; 88US-00160766.
PR 26-FEB-1988; 88US-00160771.
PR 15-JUL-1988; 88US-00219279.
PR 17-FEB-1989; 89US-00310881.
PR 05-MAY-1989; 89US-00347637.
PR 08-JUN-1989; 89US-00363138.
PR 22-OCT-1990; 90US-00600244.
PR 16-JAN-1991; 91US-00641617.
PR 26-JUL-1991; 91US-00737899.
```

PR 01-AUG-1991; 91US-00739143.
PR 31-JUL-1992; 92US-00923692.
PR 30-DEC-1992; 92US-00997733.
PR 29-DEC-1993; 93US-00176414.
PR 19-JAN-1994; 94US-00184237.
PR 14-OCT-1994; 94US-00324003.
PR 21-MAY-1999; 99US-00316572.
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX
PA (TURP/) TURPEN T H.
PA (POGU/) POGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX
PI Turpen TH, Pogue GP, Erwin RL, Grill LK;
XX
XX WPI; 2004-108227/11.
XX
XX New lysosomal enzymes, useful in treating human and animal lysosomal
PT storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
XX Example 11; SEQ ID NO 30; 71pp; English.
XX
XX The invention relates to nucleotide encoding galactosidase (GAL). The
CC invention is useful in gene therapy. The polynucleotides and polypeptides
CC are useful in treating human and animal lysosomal storage diseases, e.g.
CC Fabry's disease and Gaucher's diseases. The present sequence is Tobacco
CC mosaic virus galactosidase C-terminal peptide.
XX
XX Sequence 24 AA;
SQ
Query Match 100.0%; Score 89; DB 8; Length 24;
Best Local Similarity 100.0%; Pred. No. 7.5e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSRLRSHINPTGVLLQL 18
Db 1 TSRLRSHINPTGVLLQL 18

Search completed: December 31, 2005, 23:43:15
Job time : 7.57931 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2006, 05:47:17 ; Search time 405.497 Seconds
(without alignments)
539.100 Million cell updates/sec

Title: US-10-602-220-16

Perfect score: 2293

Sequence: 1 MQLRNELHGLCALRFLA.....EWSRLRSHINPTGTVLLQL 417

Scoring table: BLOSUM62

Xgapop 10.0	Xgapext 0.5
Xgapop 10.0	Xgapext 0.5
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 4172979 seqs, 262114271 residues

Total number of hits satisfying chosen parameters: 8345958

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US10602220/runat_30122005_140602_5569/app_query.fasta.1.782
-DB=Published Applications NA_New -QFMT=fastap -SUFFIX=p2n.rnpbn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10602220@cgn.1.1.233 @runat_30122005_140602_5569
-NCPU=6 -ICPU=3 -NO.MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DSLEXT=7

Database :

- 1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2.*
- 9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3.*
- 10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2293	100.0	1355	6	US-10-981-267-26 Sequence 26, Appl
2	2293	100.0	1547	6	US-10-981-267-25 Sequence 25, Appl
3	99	4.3	1515	7	US-11-073-112-5 Sequence 5, Appl
4	95.5	4.2	268685	6	US-10-933-025-22 Sequence 22, Appl
5	94	4.1	1241	6	US-10-750-185-56907 Sequence 56907, A
6	93	4.1	1082144	7	US-11-117-187-211 Sequence 211, App
7	91.5	4.0	2177	7	US-11-073-112-4 Sequence 4, Appl
8	90.5	3.9	2716	6	US-10-750-185-31179 Sequence 31179, A

ALIGNMENTS

RESULT 1

US-10-981-267-26
; Sequence 26, Application US/10981267
; Publication No. US20050281805A1
; GENERAL INFORMATION:
; APPLICANT: LeBowitz, Jonathan H
; APPLICANT: Beverly, Stephen
; APPLICANT: Sly, William S.
; TITLE OF INVENTION: TARGETED THERAPEUTIC PROTEINS
; FILE REFERENCE: SYM-009CP
; CURRENT APPLICATION NUMBER: US/10/981,267
; CURRENT FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/384,452
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: US 60/516,990
; PRIOR FILING DATE: 2003-11-03
; PRIOR APPLICATION NUMBER: PCT/US03/17211
; PRIOR FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: US 10/272,531
; PRIOR FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US 60/445,734
; PRIOR FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: US 60/386,019
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: US 60/408,816
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 26
; LENGTH: 1355
; TYPE: DNA
; ORGANISM: Homo sapiens

Sequence 35182, A
Sequence 36, Appl
Sequence 35, Appl
Sequence 16, Appl
Sequence 19, Appl
Sequence 8, Appl
Sequence 56545, A
Sequence 55673, A
Sequence 14, Appl
Sequence 41, Appl
Sequence 235, App
Sequence 1, Appl
Sequence 45064, A
Sequence 58314, A
Sequence 37410, A
Sequence 8553, Ap
Sequence 5, Appl
Sequence 6, Appl
Sequence 161, App
Sequence 35565, A
Sequence 45595, A
Sequence 54, Appl
Sequence 107, App
Sequence 1, Appl
Sequence 31, Appl
Sequence 8371, Ap
Sequence 61474, A
Sequence 1, Appl
Sequence 54380, A
Sequence 13307, A
Sequence 13359, A
Sequence 363, App

US-10-981-267-26

Alignment Scores:

Pred. No.: 4,55e-261 Length: 1355
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-16 (1-417) x US-10-981-267-26 (1-1355)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 48 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTGGCC 107
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 108 CTGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGACATGATGGCAAGAGCGCCT 167
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 168 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 227
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 228 GATTCTGTCATCAGTGAAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 287
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 288 TGAAGAGTACAGGTATGATGATCTCTGATGATGATGATGATGATGATGATGATGATGAT 347
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 348 GATTGAGAGGAGACTTTCAGGAGAGCCCTCAGCGCTTCTCCTCATGGATTCGCCAGCTA 407
QY 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
DB 408 GCTAAATATGTTTACAGCAAGAGACTGAAGCTAGGAGATTTATGAGATGTTGGAAATAAA 467
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
DB 468 ACTTGGCAGGCTTCCCTGGAGATTTGGATACTACAGCAATGATGCCAGACTTTGCT 527
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
DB 528 GACTGGGAGTAGACTGCTGCTAAATTTGATGATGATGATGATGATGATGATGATGATGAT 587
QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
DB 588 GCAGATGGTTATAAGCACAATGCTTGGCCCTGAAATAGGACTGGCAGAAAGCAITGTGTAC 647
QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
DB 648 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTCAAAAGCCCAATTATACAGAAATCCGA 707
QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 708 CAGTACTGCAATCACTGCGCAATTTTGTCTGACATTCATGATTCCTGGAAAGATATAAG 767
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 768 AGTATCTTGGACTGCATCTTTTAAACAGAGAGAAATGTTGATGTTGCTGGACCAAGGG 827
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 828 GGTGGATGATCCAGATATGTTAGTTGATGATGATGATGATGATGATGATGATGATGAT 887
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 888 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTTATTCATGCTTAATGACCTC 947
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320

DB 948 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGTAAGACGCTAAATTCATCAAT 1007
QY 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 1008 CAGGACCCCTTGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACTTTGAAGTGTGG 1067
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1068 GAACGACCTCTCTCAGGCTTAGCCTGGGTGTAGCTATGATAAACCGGACGAGATGGT 1127
QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1128 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGTAAAGAGCTGGCCTCTAATCCT 1187
QY 381 AlaCysPheIleThrGlnLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
DB 1188 GCCTGCTTTCATCACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGATGAGACT 1247
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
DB 1248 TCAAGGTTAAGAGTCACATAAATCCACAGGCATGTTTGTGCTTCAGCTA 1298

RESULT 2

US-10-981-267-25

; Sequence 25, Application US/10981267
; Publication No. US20050281805A1
; GENERAL INFORMATION:
; APPLICANT: LeBowitz, Jonathan H
; APPLICANT: Beverley, Stephen
; APPLICANT: Sly, William S.
; TITLE OF INVENTION: TARGETED THERAPEUTIC PROTEINS
; FILE REFERENCE: SYM-009CP
; CURRENT APPLICATION NUMBER: US/10/981,267
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/384,452
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: US 60/516,990
; PRIOR FILING DATE: 2003-11-03
; PRIOR APPLICATION NUMBER: PCT/US03/17211
; PRIOR FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: US 10/272,531
; PRIOR FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US 60/445,734
; PRIOR FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: US 60/386,019
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: US 60/408,816
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 25
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: An exemplary GILT-tagged alpha-GAL A cassette sequence

US-10-981-267-25

Alignment Scores:

Pred. No.: 5,53e-261 Length: 1547
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-16 (1-417) x US-10-981-267-25 (1-1547)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 51 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 110

```
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 111 CTCGTTTCCTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGCGCCT 170
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 171 ACCATGGGCTGGCTGCACTGGAGCGCTTATGTGCAACCTTGACTGCCAGGAGAGCCA 230
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 231 GATTTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 290
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 291 TGGAAAGATGAGGTTATAGTACCTCTGCATGTAGTACTGTTGGATGGCTCCCCAAAGA 350
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 351 GATTTCAGAGCAGACTTCAGGCAGACCTTCAGCGCTTCTCTCATGGGATTCGCAGCTA 410
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 411 GCTAAATTATGTCACAGCAAGGACTGAAGCTAGGAGTTATGCAGATGTTGGAATAAA 470
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 471 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACTACGACATTCATGCCAGACCTTTGCT 530
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 531 GACTGGGGAGTAGATCTGCTAAATTTGATGGTGTGTACTGTGACAGTTTGGAAATTTG 590
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 591 GCAGATGGTTATAAGCACATGCTCTTCCTGGCCCTGAATAGGACTGCGCAAGACATTTGTAC 650
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 651 TCCTGTGAGTGGGCTCTTTATATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 710
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 711 CAGTACTGCAATCACTGGCGAAATTTTGCCTGCAATTTGATGATTCCTGGAAAGATATAAG 770
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 771 AGTATCTTGGACTGGACATCTTTTAACGAGAGAGAAATTCGTGATGTGTGGACACGGG 830
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 831 GGTGTGGAATGACCCAGATATGTTAGTGTGATTTGGCAACTTTTGGCCTCAGCTGGAATCAGCAA 890
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 891 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 950
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 951 CGACACATCAGCCCTCAGGCAAGCTCTCTCTCAGGATNAGGACGTAATTCGCATCAAT 1010
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 1011 CAGGACCCCTTGGGCAAGCAAGGTACCACTTAGACAGGAGACAACCTTTGAAGTGTGG 1070
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1071 GAACGACCTCTCTCAGGCTTAGGCTGTAGCTATGATGATAAACCCGCGAGAGATTGGT 1130
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1131 GGAACCTCGCTCTTATACCATCGCAGTGTGCTTCCCTGGGTAAAGAGTGGCCTGTATCT 1190
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
```

```
Db 1191 GCCTGCTTCATCATCACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGATGGACT 1250
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1251 TCAGGTTAAGAAGTACATAAATCCACAGGCACTGTTTGTCTTACAGTA 1301
RESULT 3
US-11-073-112-5
; Sequence 5, Application US/11073112
; Publication No. US20050260627A1
; GENERAL INFORMATION:
; APPLICANT: Hintz et al.
; TITLE OF INVENTION: Mannosidases and Methods for using the Same
; FILE REFERENCE: 62447-02
; CURRENT APPLICATION NUMBER: US/11/073,112
; PRIOR FILING DATE: 2005-03-03
; PRIOR APPLICATION NUMBER: US 10/089,211
; PRIOR FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: PCT/US00/27210
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/157,341
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 5
; LENGTH: 1515
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-11-073-112-5
Alignment Scores:
Pred. No.: 0.484 Length: 1515
Score: 99.00 Matches: 62
Percent Similarity: 35.74% Conservative: 32
Best Local Similarity: 23.57% Mismatches: 74
Query Match: 4.32% Indels: 95
DB: Gaps: 18
US-10-602-220-16 (1-417) x US-11-073-112-5 (1-1515)
QY 27 ProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrProThrMetGlyTrpLeuHis 46
Db 669 CCCGAGCCAGCCAGTCACTA-----CGAACCGTTCCCTCGATTGAT 707
QY 47 TrpGluArgPheMet---CysAsnLeuAspCysGlnGluProAspSerCysIleSer 65
Db 708 TGGAAAGCAGTCAATATTGCCGACGCAAGCTCGCAA-----TGTCTCA 752
QY 66 GluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLys-AspAlaG1 85
Db 753 CATCAGCTG-----GAATGGTGGCGCAGACTCGTA 782
QY 85 YTrpGluTyrLeu-----CysIleAspAs 93
Db 783 CTACGAGTACCTGATCAAGATGTACTACGATCCGCAACGCTTTGGCCTCTACCGGA 842
QY 93 pCysTrpMetAlaProGlnArgAspSerGluGlyArgLeuGlnAlaAspProGlnArgPh 113
Db 843 CGCTGGTGGTGGAGCTGGCGAGTCGAGCATCAACCATCTGGCTTCGCACCCGCCACCCG 902
QY 113 eProHisGlyIleArgGlnLeuAlaAsnTyrValHisSerLysGlyLeuLysLeuGly11 133
Db 903 CCCA---GACGTGACTTTCTTGGCCACTTAC-----AACGAGGAGCATCAGCTGGGCGCT 953
QY 133 eTyrAlaAspValGlyAsnLysThrCysAlaGlyPheProGlySerPhe-----GlyTy 151
Db 954 GACCAAGCCAA-----CACTGACCTGCTTCGAC---GGTGGAGAGCTTTCTGCTGTGG 1004
QY 151 rTyrAspIleAspAlaGlnThrPheAlaAspTrpGlyValAspLeuLeuLysPheAspG1 171
Db 1005 GACATTGCTGGAGCCGCGAGGACTTTGTCGACTTTCGCGCTTGACCTTGTGCC-----GG 1058
```

QY 171 yCys-----TyrCysAspSerLeuGluAsnLeuAla----- 181
 DB 1059 GTGCCACGAGAGCTCAACTCGACTCTGACGGGCATCGGCCCTGACCAATTGAGCTGGGA 1118
 QY 182 -----AspGlyTyrIlyshiMetSerLeuAlaLeuAsnArgThrGly-- 195
 DB 1119 CCTAACGGGTGTGCCGACAGCCAGAGAGCTG-----TTCCAGCGCGCAGGCTT 1169
 QY 196 -----ArgSerIleValTyrSe 201
 DB 1170 CTACATCAACAGCGGCCAATACATTCTGTCGCCGAAGTATCGAGAGCTTCTACTATGC 1229
 QY 201 r-----CysGluTyrProLeuTyrMetTyrProPheG1 212
 DB 1230 ATGGCGCGTACAGGTGATGGAACTGACCTCGAATGG-----GTGTGG----- 1272
 QY 212 nlyAsProAsnTyrThrGluIleArgGlnTyrCysAsnHisTyrArgAsnPheAlaAsp11 232
 DB 1273 ----AAGCGCTTCACCAACATCAACAAGTACTGCCCATCGCACCGGTTTCGCGGGCT 1328
 QY 232 eAspAsp 234
 DB 1329 GGAGAAC 1335

RESULT 4

US-10-933-025-22/c
 ; Sequence 22, Application US/10933025
 ; Publication No. US20050265987A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ROSEN, STEVEN
 ; APPLICANT: HEMMERICH, STEFAN
 ; APPLICANT: TOMITA, MEGUMI
 ; TITLE OF INVENTION: Sulfotransferases and methods of use
 ; TITLE OF INVENTION: thereof
 ; FILE REFERENCE: UCAL-230CON
 ; CURRENT APPLICATION NUMBER: US/10/933,025
 ; CURRENT FILING DATE: 2004-09-01
 ; PRIOR APPLICATION NUMBER: 10/025,966
 ; PRIOR FILING DATE: 2001-12-21
 ; PRIOR APPLICATION NUMBER: 60/258,577
 ; PRIOR FILING DATE: 2000-12-27
 ; PRIOR APPLICATION NUMBER: 60/267,831
 ; PRIOR FILING DATE: 2001-09-02
 ; NUMBER OF SEQ ID NOS: 26
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 22
 ; LENGTH: 268685
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)..(268685)
 ; OTHER INFORMATION: n = A, T, C or G
 US-10-933-025-22

Alignment Scores:

Pred. No.: 2,48e+03 Length: 268685
 Score: 95.50 Matches: 87
 Percent Similarity: 30.73% Conservative: 39
 Best Local Similarity: 21.22% Mismatches: 121
 Query Match: 4.16% Indels: 163
 DB: 6 Gaps: 22

US-10-602-220-16 (1-417) x US-10-933-025-22 (1-268685)

QY 80 GlyTyrPlyAspAlaGlyTyrGluTyr---LeuCysIleAspAspCysTyrMet----- 96
 DB 57945 GGGTGGGGGATGGAGATGGGAATCGGGGTGTAGCAGGAGTCACTGGGTGTGAGGA 57886
 QY 97 -----AlaProGlnArgAspSerGluGly-----Arg 105
 DB 57885 GGCCACGCTGGCCCTCGATGCGCCAGAGTCAGGTATGTGAGGCTCACTGTTCAGAGG 57826

QY 106 LeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeuAlaAsnTyrValHis 125
 DB 57825 AGTGAAGACACCCCAAGA-----ACACAT 57799
 QY 126 SerIysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLysThrCysAlaGlyPhe 145
 DB 57798 AGTAAG-----GCTAACACAGGGGAGAGAGGAGGCCAGGGGCTTT 57760
 QY 146 ProGlySerPheGlyTyr-----TyrAspIleAspAlaGlnThrPheAlaAspTyr 162
 DB 57759 ---GACTCTTTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 57703
 QY 163 GlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeuAlaAsp 182
 DB 57702 GGAGTG-----CAGGGGTGTGATCTCGGCTCCTCGCACTCGCACTTCC 57658
 QY 183 GlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyrSerCys 202
 DB 57657 TGGTTCAAGCAATTTCTGCCTCAGCCTC-----CCAAGT 57625
 QY 203 GluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArgGlnTyr 222
 DB 57624 AGCTGGGATTACAGGCACAGCCACCCAGCTAAT-----TTT 57583
 QY 223 CysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLysSerIle 242
 DB 57582 TGTATTTTATTAGATAT-----GGGGTTTCGCCATGTGG----- 57547
 QY 243 LeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGlyGlyTyr 262
 DB 57546 -----CCAGGCTGGTCT 57535
 QY 263 AsnAspProAspMetLeuValIleGlyAsnPheGlyLeu----- 275
 DB 57534 CAAACTCCTGACCTCAAGTGATCTGCCACCTCGGCTCCCAAGTACTCGGGTTACAGG 57475
 QY 276 -----SerTyrAsnGlnGlnValThrGln 283
 DB 57474 CGTGAGCCACCGCCCGGCTCTACTCTTTTCCACCTTGGCGGAGCTCCCTGAGCCCT 57415
 QY 284 MetAlaLeuTyrAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeuArgHisIle 303
 DB 57414 GTACAACCTTGGACAGAGGCTGGGGCCAGCTTGGTCTCTG-----ATC 57373
 QY 304 SerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsnGlnAspPro 323
 DB 57372 AGCACCCAGGGCGGCTCTCTGCTGGGCTCTGCTGGAGCATTTCTTG-----CCT 57322
 QY 324 LeuGlyLysGlnGlyTyrGlnLeu---ArgGlnGlyAspAsnPheGluValTyr----- 340
 DB 57321 GCCTTCCAGCAAGACCTTGGCTTGGCCAGGTGAGTCTGACCTGAGCAGTGTCTTCTGTCCG 57262
 QY 341 -----GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGlu 358
 DB 57261 TCACAGCCTGGTCTCTGCCAGGACTTGTGTAATCGCTGCCACTTG----- 57214
 QY 359 IleGlyGlyProArgSerTyr----- 365
 DB 57213 -----GCATCCAGGTCTGTCTCCACCTCGGATGGCTCTTGGGTGACTCATCAATTGT 57160
 QY 366 -----ThrIleAlaValAlaSerLeuGly-----LysGlyValAlaCys 378
 DB 57159 GCTTATATCCGTCGTCTGTGGCTCTGCGAGGTCTGTGATGTAATCCAGGGCTGGCTGCC 57100
 QY 379 AsnProAlaCysPhe----- 383
 DB 57099 AACCGAGCTTGCACAGCAGGCGCTCTGGGGTGAGCCTGTGGCTTATCTAGTCGGTCCCTTT 57040
 QY 384 -----IleThrGlnLeuProValLysArgLysLeuGly 395
 DB 57039 CACCTCAGAGCGAGTGTATTTATGATGATACAGCTACTGTGTTGAATATTCACTGCAC 56980
 QY 396 PheTyrGluTyrThrSerArgLeuArgSer 405


```

Db      56979 ATATAT---GGGACGTGAGTCTGAGTAGC 56953
RESULT 5
US-10-750-185-56907
; Sequence 56907, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; FILE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56907
; LENGTH: 1241
; TYPE: DNA
; ORGANISM: Bovine 19866881771060
US-10-750-185-56907

Alignment Scores:
Pred. No.:      1.41      Length:      1241
Score:          94.00     Matches:      92
Percent Similarity: 30.51%  Conservatives: 45
Best Local Similarity: 20.49% Mismatches:    132
Query Match:      4.10%   Indels:       180
DB:               6      Gaps:         21

US-10-602-220-16 (1-417) x US-10-750-185-56907 (1-1241)
QY      6 ProGluLeuHisLeuGlyCysAlaLeuAlaLeuAraGluLeuAlaLeuValSerTrpAsp 25
Db      226 CCAATTTCTGGAATATCTCTGCTACATGTTCTTTCCACATTTCCGAGGCAAGTGGGAA 285
QY      26 IleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrProThrMetGlyTrpLeu 45
Db      286 -----ACC 288
QY      46 HisTrpGluArgPheMetCysAsnLeuAspCysGlnGluProAspSer-Cystile-- 64
Db      289 CACTGTGATGAGCAATATGCTCTCTCCAAAGTCCAGGAGAGGTGGGAGCAAAACGGTGA 348
QY      65 -----SerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLys 82
Db      349 GCCTGGGAACGGGGTGTGTTCTGCGCCCTTGAGGAA-----TGGGG 390
QY      82 sAspAlaGlyTyxGluTyxLeuCysIleAspAspCysTrpMetAlaProGlnArgAspse 102
Db      391 AGATCCA-----GCTCCTAAGCAGCAAG 414
QY      102 rGlu-----Glyar 105
Db      415 GGAGTTTCTCGGTGTCCCCCTGCCAAAGCACTGGGAGGTGGGACAGGTTGAAACTCCAA 474
QY      105 gLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeuAlaAsnTyxValHi 125
Db      475 GTCTCAGGCCCTGGCAGGGAATTTCCACAGCAGTT---GGCTGATTTGGATACACTCA 531
QY      125 sSerLysGlyLeuLysLeuGlyIleTyxAlaAspValGlyAsnLysThrCysAlaGlyPh 145
Db      532 TGCTAAGTCCCTC-----TGAGAGCTAGAATTTTTCACITTCCTTTGGGAT 576
QY      145 eProGlySerPheGlyTyxTrpAspIleAspAlaGlnThrPheAlaAspTrpGlyValas 165

```

RESULT 6

```

US-11-117-187-211
; Sequence 211, Application US/11117187
; Publication No. US20050266560A1
; GENERAL INFORMATION:
; APPLICANT: PREUSS, DAPHNE
; APPLICANT: COPENHAVER, GREGORY
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS
; FILE REFERENCE: ARCD:309US
; CURRENT APPLICATION NUMBER: US/11/117,187
; CURRENT FILING DATE: 2005-04-28
; PRIOR APPLICATION NUMBER: US/09/531,120
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/125,219
; PRIOR FILING DATE: 1999-03-18

```

```
/ NUMBER OF SEQ ID NOS: 212
/ SOFTWARE: Patent In Ver. 2.1
/ SEQ ID NO 211
/ LENGTH: 1082144
/ TYPE: DNA
/ ORGANISM: Arabidopsis thaliana
US-11-117-187-211

Alignment Scores:
Pred. No.: 3,78e+04 Length: 1082144
Score: 93.00 Matches: 76
Percent Similarity: 30.97% Conservatives: 42
Best Local Similarity: 19.95% Mismatches: 117
Query Match: 4.06% Indels: 146
DB: 7 Gaps: 20

US-10-602-220-16 (1-417) x US-11-117-187-211 (1-1082144)
QY 3 LeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAlaLeuVal 22
Db 565984 CTGCGTTATCCGAGGTTAGA---GCGGCGCGCGCGCGCTTAGAATAGTCGGA----- 566034
QY 23 SerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrProThrMet 42
Db 566035 -----GATCTTCGGAACCTCGGTTGCTGAT-----TTACACCTCGCATG 566076
QY 43 GlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluProAspSer 62
Db 566077 GACTTGATGAGCTGGCTGCTCTCTCTTTGGTTTCAAAATCGATAACGTTGGAATCAG 566136
QY 63 CysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLys 82
Db 566137 -----CGAGAGAATCTCGTTCTTCACTTAGCT----- 566163
QY 83 AspAlaGlyTyrrGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArgAspSer 102
Db 566164 -----AACTCT 566169
QY 103 GluGlyArgLeuGlnAlaAspProGlnArgPheProHisGly----- 116
Db 566170 CAAATCGCGTGCA--CCTCGCCACGATCATCCGCGGTTCGATCGAGCTGTTCTC 566226
QY 117 -----IleArgGlnLeuAlaAsnTyrrValHisSerLysGlyLeu 129
Db 566227 CGGAGATTTCCGAAGAAGCTGCTTCTGTAATATACAAATTTGGTCTGTTCTCGGAGTG 566286
QY 130 LysLeuGlyIleTyrrAlaAspValGlyAsnLysThrCysAlaGlyPheProGlySerPhe 149
Db 566287 AGGTGCCACGTACCGCCCTATCCAAAGCCGT----- 566319
QY 150 GlyTyrrTyrrAspIleAspAlaGlnThrPheAla----- 160
Db 566320 -----CACCAAAAGCAAGCGGTTTGAATCTCGCCGCGAGCTTCTC 566361
QY 161 -----AspTrpGly-----ValAspLeuLeuLysPheAspGly 171
Db 566362 TATGTTGCTCTCTATCTATTGATTTGGGGAGAATCTCGGAATCTCTGTTTATGCGCTGAG 566421
QY 172 CysTyrrCysAspSerLeuGluAsnLeuAlaAspGlyTyrrLysHisMetSerLeuAlaLeu 191
Db 566422 TGCTGTGTTACATT-----TTTCATCATATGGCTATGGAACCTT 566460
QY 192 AsnArg-----ThrGlyArgSerIleValTyrrSerCys 202
Db 566461 AATAAAGTCTCGCGGAGAGTTTGATGATATGACTGGAATGCTTAC----- 566508
QY 203 GluTrpProLeuTyrr-----MetTrpProPheGln 212
Db 566509 ---TGGCTTCTCTCAGTGATGTTGCTTTCTCAAGAGTGTGGTCACTGCTTATTAC 566565
QY 213 LysProAsnTyrrThrGluIleArgGlnTyrrCysAsn-----HisTrp 226
Db 566566 AAGACGTTTAAGACGGAGGTTGAGTAGCAACATGGGACGAGCCACACTCTGCTTGG 566625
```

```
QY 227 ArgAsnPheAlaAspIleAspAsp-----SerTrpLysSerIleLys 240
Db 566626 AGAAATTACGACGACATCAATGAGTATTTCTGGAGTAAGAGAGCTTGAAGAGCCTTAA 566685
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 566686 TGGCCTCTTGACTACACCACT-----AATTTCTTTGATACTACA----- 566724
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeu----- 275
Db 566725 -----CCTAAAAGCAGTAGAGTGGGGAACAGAGGATTTGTGGAGCAGAGATCA 566772
QY 276 SerTrpAsnGlnGlnValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPhe 295
Db 566773 TTTTGGAAATGTGTATCGAAGCTTTTGATCGGTTATGGATTTATTCGTT-----CTGTAC 566826
QY 296 MetSerAsnAspLeuArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAsp 315
Db 566827 TTTCAGGACGCTATTATTGTTGCTACTAGTATGTTAAGTTTCTTGGCAAGACCGTGAT 566886
QY 316 Val 316
Db 566887 GTA 566889

RESULT 7
US-11-073-112-4
/ Sequence 4, Application US/11073112
/ Publication NO. US20050260627A1
/ GENERAL INFORMATION:
/ APPLICANT: Hintz et al.
/ TITLE OF INVENTION: Mannosidases and Methods for using the Same
/ FILE REFERENCE: 62447-02
/ CURRENT APPLICATION NUMBER: US/11/073,112
/ CURRENT FILING DATE: 2005-03-03
/ PRIOR APPLICATION NUMBER: US 10/089,211
/ PRIOR FILING DATE: 2002-02-25
/ PRIOR APPLICATION NUMBER: PCT/US00/27210
/ PRIOR FILING DATE: 2000-10-02
/ PRIOR APPLICATION NUMBER: 60/157,341
/ PRIOR FILING DATE: 1999-10-01
/ NUMBER OF SEQ ID NOS: 25
/ SOFTWARE: Patent In version 3.3
/ SEQ ID NO 4
/ LENGTH: 2177
/ TYPE: DNA
/ ORGANISM: Aspergillus nidulans
US-11-073-112-4

Alignment Scores:
Pred. No.: 6.37 Length: 2177
Score: 91.50 Matches: 45
Percent Similarity: 41.72% Conservatives: 23
Best Local Similarity: 27.61% Mismatches: 49
Query Match: 3.99% Indels: 46
DB: 7 Gaps: 11

US-10-602-220-16 (1-417) x US-11-073-112-4 (1-2177)
QY 27 ProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrProThrMetGlyTrpLeuHis 46
Db 1059 CCCGGAGCCAGCCAGTA-----CGAACCGTTCCTCGATTTGAT 1097
QY 47 TrpGluArgPheMet---CysAsnLeuAspCysGlnGluProAspSerCysIleSer 65
Db 1098 TGGAGCGCAGTCAATATTGCGGCAAGCTCGCCAA-----TGTCA 1142
QY 66 GluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLys-AspAlaGly 85
Db 1143 CATCAGCTG-----GAATGCTGCGCAGACTCGTA 1172
QY 85 TyrrGluTyrrLeu-----CysIleAspAsp 93
```

Db 1173 CTACGAGTACGTCAAGATGTACGTCTACGATCCGACCGCTTTGGCTCTACCGGGA 1232
QY 93 pCysTrpMetAlaProGlnArgAspSerGluGlyArgLeuGlnAlaAspProGlnArgPh 113
Db 1233 CCGCTGGGTGCGAGTCCGAGTCCGAGTCAACCATCTGGCTTCGACCCGTCACCCG 1292
QY 113 eProHisGlyIleArgGlnLeuAlaAsnTyrValHisSerLysGlyLeuLeuGlyIle 133
Db 1293 CCCA---GACGTGACTTCTTGGCCACTTAC-----AAGAGGAGCATCAGCTGGGCC 1343
QY 133 eTyrAlaAspValGlyAsnLysThrCysAlaGlyPheProGlySerPhe-----GlyTyr 151
Db 1344 GACCAGCCAA-----CACTGACCTGCTTCGAC---GGTGAAGCTTCTGCTGGTGG 1394
QY 151 rTyrAspIleAspAlaGlnThrPheAlaAspTrpGlyValAspLeuLysPheAspG 171
Db 1395 GACATTGCTGACCGCCAGGACTTGTGCACTTTCGGCTTGCCTTGCGCC-----GG 1448
QY 171 yCysTyr 173
Db 1449 CTGCCAC 1455

RESULT 8

US-10-750-185-31179
; Sequence 31179, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31179
; LENGTH: 2716
; TYPE: DNA
; ORGANISM: Bovine 19866880843346
US-10-750-185-31179

Alignment Scores:
Pred. No.: 11.6 Length: 2716
Score: 90.50 Matches: 48
Percent Similarity: 30.84% Conservatives: 22
Best Local Similarity: 21.15% Mismatches: 82
Query Match: 3.95% Indels: 76
DB: 6 Gaps: 11

US-10-602-220-16 (1-417) x US-10-750-185-31179 (1-2716)

QY 38 ArgThrProThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCys--- 56
Db 1674 AGTTCCCGAGCTCGGAGCCCTCAGCTGTAAAGATC-----CACCTGAATCGGG 1727
QY 57 GlnGluProAspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMet 76
Db 1728 AGACCTGAGTTCGATCCCTGCGTGGGAAG----- 1757
QY 77 ValSerGluGlyTrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMet 96
Db 1758 -----ATCCCTCGGAGAGGGGAGGCTCTCCACTCTGCGCTGGAGAAATTTTCATGGACT 1811
QY 97 AlaProGlnArgAspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGly 116
Db 1812 GTATAGTCCATAGGGTCCGAGAGTCGGACACAA-----CCAAGCGACTTT----- 1856

QY 117 IleArgGlnLeuAlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAsp 136
Db 1857 -----CACTTCATTCGGGACAAAGGGCTGTATTGGCGTT----- 1892
QY 137 ValGlyAsnLysThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAla 156
Db 1893 -----TGTACCAAGTTCCTCTGTCTTTGGCTGGGCAAGGCGGATGCT 1937
QY 157 GlnThrPheAlaAspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSer 176
Db 1938 ACCTCTGTTGGCTC-TGG-----TCAGCTGT 1963
QY 177 LeuGluAsnLeuAlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArg 196
Db 1964 GTCCAGAGTCTGGCTGATGGA-----GGTGCTCAGGTGGGACGTGAGGACTG 2011
QY 197 SerIleValTyrSerCys-----GluTyrProLeu 206
Db 2012 GGAGTGGCCAGGGCTGTGGCCAGGGCCCTGACAGCCTCTGCCACTCTACTGACACAG 2071
QY 207 TyrMetTrpProPheGlnLys-----ProAsnTyr 216
Db 2072 GCATGTGTGCGCCAGCAGTCCCTAGGAGTGTGCTCTGCCAAGCCCTCTTGTCCACGCTC 2131
QY 217 ThrGluIleArgGlnTyr-----CysAsnHisTrpArgAsnPhe 229
Db 2132 ACGTCTCTCCAGGCCCAAGAGCTGCACACCCTCTCTGTGCTGCCCTGGGGAGACCTT 2191
QY 230 AlaAspIleAspAspSerTrp 236
Db 2192 TCCCACTGGGACCAAGTCATGG 2212

RESULT 9

US-10-750-185-35182
; Sequence 35182, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35182
; LENGTH: 1766
; TYPE: DNA
; ORGANISM: Bovine 19866880928175
US-10-750-185-35182

Alignment Scores:
Pred. No.: 8.09 Length: 1766
Score: 89.50 Matches: 57
Percent Similarity: 31.33% Conservatives: 21
Best Local Similarity: 22.89% Mismatches: 85
Query Match: 3.90% Indels: 86
DB: 6 Gaps: 15

US-10-602-220-16 (1-417) x US-10-750-185-35182 (1-1766)

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAla----- 37
Db 481 CTCTGTCCATGGGATTTCCAGGCAAG-----AATACTGGAATCAGTTGCCATTTC 531

```
QY 38 -----ArgThrProThrMetGly-----TrrLeuHisTrpGlu-Ar 49
Db 532 CTTCTCCAGGAGATCTTGCCAAACCAAGATCGAACCCATTCTCTTCACTGGGAGCG 591
QY 49 gPheMetCysAanLeuAapCysGlnGluProAaspSerCysIleSerGluLysLeuPh 69
Db 592 GATTCTTCACTGAGCCAGCCAGGAGCCCACTAGTGTGTAGCACATTTTATA 651
QY 69 eMetGluMetAlaGluLeuMetValSerGluGlyTrpLysAap-----AlaG1 85
Db 652 CCTTCAA-----GAGGAAAGGTGGGAGGAAACAGTCTTGGGGGCGAG 693
QY 85 yTyrGluTyrLeu-----CysIleAasp-----CysTrpMetAlaProGlnArgAaspSe 102
Db 694 ATATGCCCTCTGCTCTGTTCTTCTCTCTGTTCTGAGGCTGGAGCATCAGG----- 748
QY 102 rGluGlyArgLeuGlnAlaAaspProGlnArgPheProHisGlyIleArgGlnLeuAlaAs 122
Db 749 -GAGGGATGCTGGTGCCTCT----- 769
QY 122 nTyrValHisSerLysGlyLeuLysGlyIleTyrAlaAaspValGlyAsnLysThrCy 142
Db 770 -----GATGTCAAGAGCCCAAGCTTGCCCACTTAAGTGCAGGGGAACTGCAGACTG 822
QY 142 salaglyPheProGlySerPheGlyTyrTyrAaspIleAaspAlaGlnThrPheAlaAaspTr 162
Db 823 CCAGGGGCTGACAGAGTCAATCGGGA-----GGCTG 852
QY 162 pGlyValAaspLeuLys-----PheAaspGlyCysTyrCysAs 175
Db 853 GAGAACTGGCTTACAAAGACGAGGAGCAATCTGAGGCTGGGATGCACAATGGGCAGA 912
QY 175 pSerLeu-----GluAsnLeuAlaAaspG1 183
Db 913 TGCTCTAGGAGGAGGAGGCTGGGTAGTGGCTGACTTCTTAGGAGAGGAGGCACTGG 972
QY 183 yTyrLysHisMetSerLeuAlaLeuAsnArgThr-----GlyArgSerIleVa 199
Db 973 CTATGTCCAGTGTGCATCACCGACCCAGGATTTTAGTTCAGGGGGAGAGATCCAGT 1032
QY 199 lTyrSerCys---GluTyrProLeuTyrMetTrpProPheGlnLysProAsnTyrThrG1 218
Db 1033 TGCTCTCTGTGTCTCTGGCTGTC-----CCTTCTCTCTACCA 1071
QY 218 uIleArgGlnTyrCysAsnHisTrp 226
Db 1072 GACTGAACCTGTAGGAGGACACTGG 1096
```

RESULT 10

```
US-11-112-908-36
; Sequence 36, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 36
; LENGTH: 98345
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (49350)..(49350)
; OTHER INFORMATION: n is a, c, g, or t
US-11-112-908-36

Alignment Scores:
Pred. No.: 4,4e+03 Length: 98345
Score: 88.00 Matches: 49
Percent Similarity: 32.43% Conservatives: 15
Best Local Similarity: 24.81% Mismatches: 75
Query Match: 3.84% Indels: 58
DB: 7 Gaps: 11

US-10-602-220-16 (1-417) x US-11-112-908-36 (1-98345)

QY 225 HisTrpArgAsnPheAlaAaspIleAaspSerTrpLysSerIleLysSerIleLeuAasp 244
Db 76964 CATTGCCGA----- 76972
QY 245 TrpThrSerPheAsnGlnGluArgIleValAaspValAlaGlyProGlyGlyTrpAanAasp 264
Db 76973 -----CAGAGCGGGACAGCGCCGCTCGGGCCCGAGAGGG----- 77008
QY 265 ProAaspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsn-----GlnGlnValThr 282
Db 77009 ---GAGCGGTACAGCGGGGCGCTGGCACCGCGGTGGAACTCGGCGCGGGGTCC 77065
QY 283 GlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAaspLeuArgHis 302
Db 77066 GCCCTGTGGCCCTGGCGCCCGCCGCTGCTGCTTGTCTCTCG----- 77110
QY 303 lIeSerProGlnAlaLysAlaLeuLeuGlnAaspLysAaspValIleAlaIleAanGlnAasp 322
Db 77111 CTCGCCCTCTCCCGCCCGCTCGCTTCAGCGCGCGCGAGGAGCTAACCGCGGT 77170
QY 323 ProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAaspAsnPheGluValTrpGluArg 342
Db 77171 CTTCTAGGAACCTCGGGCGGGCAGCAGCCCGCGGA-----TTCTGTGGCTGCT 77221
QY 343 ProLeu-----SerGlyLeuAlaTrpAlaValAla-----MetIleAsnArgGlnGluIle 359
Db 77222 CGCTGCCATGAAGCGGGACCGCTGGCGCGCTTCTGTCTCTGGGTGCTCCCGACAGT 77281
QY 360 GlyGlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsn 379
Db 77282 GCGGGGCTCTCG-----ACGCGCGCGCGCGGTACGCGGACTCGGGCGCGC----- 77329
QY 380 ProAlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrp 399
Db 77330 -----CTTCCCTTAGCGTGGCGAGGG-----TGG 77356
QY 400 ThrSerArgLeuArg-----SerHisIleAsnProThrGlyThr 412
Db 77357 ACGGGCGAGCGCGCGCGCTGGGGCGCGGTCTGGGGTCTCTCGGGGACG 77407

RESULT 11
US-11-112-908-35
; Sequence 35, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 36
; LENGTH: 98345
; TYPE: DNA
; ORGANISM: Homo sapiens
```

; FILE REFERENCE: 054163-5003-US
; CURRENT APPLICATION NUMBER: US/11/038,284
; PRIOR FILING DATE: 2005-01-21
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 35
; LENGTH: 127340
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-35

Alignment Scores:
Pred. No.: 6.43e+03 Length: 127340
Score: 88.00 Matches: 49
Percent Similarity: 32.49% Conservatives: 15
Best Local Similarity: 24.87% Mismatches: 75
Query Match: 3.84% Indels: 58
DB: 7 Gaps: 11

US-10-602-220-16 (1-417) x US-11-112-908-35 (1-127340)

```
QY 225 HieTTPArgAsnPheAlaAspSerTrpLysSerIleLeuAsp 244
|||||
Db 25035 CATGGCGA----- 25043
QY 245 TrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGlyGlyTrpAsnAsp 264
|||||
Db 25044 -----CAGAGCGGGCAGCGCCCTCGGGCCCGAGAGGG----- 25079
QY 265 ProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsn-----GlnGlnValThr 282
|||||
Db 25080 ---GAGCGGTACAGCGGGGGCGCTGGCAGCGGGTGGAACTCGCGCGGGGTCC 25136
QY 283 GlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeuArgHis 302
|||||
Db 25137 GCCCTGGCCCTGGCCCGCCCGCTCGCTTTGCTTCTCG----- 25181
QY 303 IleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaLeuAsnGlnAsp 322
|||||
Db 25182 CTCGGCCCTCCCGCCCGCCCGCTCGCTTCCAGCGCGCGAGCGCTTAACCGCGGGT 25241
QY 323 ProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrpGluArg 342
|||||
Db 25242 CCTCTAGGACCTCGGGCGGGGAGCAGCCCGCGGA-----TTCTGCTGGCTCCT 25292
QY 343 ProLeu-----SerGlyLeuAlaTrpAlaValAla-----MetIleAsnArgGlnGluIle 359
|||||
Db 25293 CGCTGCCATGAAGCGGACCGCTGGCGCGCTTCTGCTCTCTGCTGCTCCGACAGT 25352
QY 360 GlyGlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsn 379
|||||
Db 25353 GCGGGGCTCGG-----ACGGCGGGCGGGCGCTCAGCGGACTCGGGGCGGCGC----- 25400
QY 380 ProAlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrp 399
|||||
Db 25401 -----CTTCCCTTAGCGGTGGCGCGAGGG-----TGG 25427
QY 400 ThrSerArgLeuArg-----SerHisIleAsnProThrGlyThr 412
|||||
Db 25428 ACGGGCGAGCGGCGCGCGCTGGGGCCCGGTGGGGTCTCTCGGGGAGC 25478
```

RESULT 12

US-11-038-284-16
; Sequence 16, Application US/11038284
; Publication No. US20050246793A1
; GENERAL INFORMATION:
; APPLICANT: COOKE, DAVID
; APPLICANT: DEBET, MARTINE
; APPLICANT: GIDLEY, MICHAEL, JOHN
; APPLICANT: JOBLING, STEPHEN, ALAN
; APPLICANT: SAFFORD, RICHARD
; APPLICANT: SIDEBOTTOM, CHRISTOPHER, MICHAEL
; APPLICANT: WESTCOTT, ROGER, JOHN
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO PLANT STARCH COMPOSITION

; FILE REFERENCE: 054163-5003-US
; CURRENT APPLICATION NUMBER: US/11/038,284
; PRIOR FILING DATE: 2005-01-21
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 35
; LENGTH: 127340
; TYPE: DNA
; ORGANISM: Solanum tuberosum
US-11-038-284-16

Alignment Scores:
Pred. No.: 24.3 Length: 2576
Score: 87.50 Matches: 43
Percent Similarity: 37.14% Conservatives: 22
Best Local Similarity: 24.57% Mismatches: 49
Query Match: 3.82% Indels: 61
DB: 9 Gaps: 9

US-10-602-220-16 (1-417) x US-11-038-284-16 (1-2576)

```
QY 108 AlaAspProGlnArgPhe-----ProHisGlyIleArgGlnLeuAlaAsnTyrValHis 125
|||||
Db 1138 GCACCAAGCAGCGGTTTGGACGCCGACGACCTTAAAGTCTTTGATTGATAAGCTCAT 1197
QY 126 SerLysGlyLeuLysLeu----- 131
|||||
Db 1198 GAGCTAGGAATTGTTGTTCTCATGGACATGTTTCAGCCATGCATCAATAATACTTTA 1257
QY 132 ---GlyIleTyrAlaAspValGlyAsnLysTyrCysAlaGlyPheProGlySerPheGly 150
|||||
Db 1258 GATGGACTGAACATGTTTGACGCCACCGATAGTGTACTTCACTCTGGAGCTCGTGGT 1317
QY 151 TyrTyrAspIle---AspAlaGlnThrPhe----- 159
|||||
Db 1318 TATCATGTGATGGGATTCGCCGCTTTTAACTATGGAACTGGGAGGTACTTAGGTAT 1377
QY 160 -----AlaAspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAsp 175
|||||
Db 1378 CTCTCTCAAAATGCGAGATGGTGGTGGATGAGTTCAAATTTGATGATTTAGATTGAT 1437
QY 176 SerLeuGluAsnLeuAlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGly 195
|||||
Db 1438 GGTGTGACATCAATGATG-----TATACTCACCGGATTTATCGGTGGGATTCACCTGG 1491
QY 196 Arg-----SerIleValTyrSer-CysG1 203
|||||
Db 1492 AACTACAGGAATACATTTGGACTCGCAACTGATGTGGATCTGTGTGTATCTGATGCTG 1551
QY 203 utrProLeu-TyrMetTrpProPheGlnLysProAsnTyrThrGluIleArgGlnTyrC 223
|||||
Db 1552 GTCAAGCATCTTATTCATGGGCTTTTCCAGATGCAATTAC----- 1592
```

RESULT 13

US-11-038-284-19
; Sequence 19, Application US/11038284
; Publication No. US20050246793A1
; GENERAL INFORMATION:
; APPLICANT: COOKE, DAVID
; APPLICANT: DEBET, MARTINE
; APPLICANT: GIDLEY, MICHAEL, JOHN


```
Qy 197 rle-----ValTyrSerCysGluTrpPro----- 205
Db 2515 ACTTAATAGATCTTCAATATGCTTGGCTCAACCGTGAAGTTATAGCAGATGACTTGC 2574
Qy 206 -----LeuTyrMetTrpProPheGlnLysProAs 215
Db 2575 CTCATCTCTTGGCTTAACCTAGTACCAAGATTATATCTACTGCAGCGACTGGAGCCGAC 2634
Qy 215 nTyrThrGluLeuArgGlnTyrCysAsnHisTrp-ArgAsnPheAlaAspIleAspAsps 235
Db 2635 GCAGCATTTAGCGTCCCAACA-----ACCAGTGGCCAAACCGCACCATTCATTCAGGGCC 2691
Qy 235 eTrpLysSerIleLysSerIleLeuAspTrpThrSerPheAsnGlnGluArgIleVala 255
Db 2692 ATTTGGATTATGTATGATGCATCTCTCTTCACTCATCTCGACAGTCAAGGTTGGAATG 2751
Qy 255 spValAlaGlyProGlyGlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyL 275
Db 2752 AATGTCTTCCAGCAATGGGCATGCTC-CCACCTCTGTGGCTGTGGCAGTTGGGGG- 2809
Qy 275 euSerTrpAsnGlnValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuP 295
Db 2810 -----TTTGTGTGTG-----ATGCCCTGCCACTAC 2837
Qy 295 he-MetSerAsnAspLeuArgHisIleSerProGlnAlaLysAlaLeuLeu----- 311
Db 2838 TCTCTTAATGCTGACACAGAGACTTGTAGTGTCTCTACGACTTTCCTGTCTTCAAGTCAA 2897
Qy 312 -----GlnAspLysAspValIle----- 317
Db 2898 AAGAGTGGCCATCAACCGCATGGTGTATGTAAACAACAGAGCCCGCATCTCTTCC 2957
Qy 318 -----AlaIleAsnGlnAspProLeuGlyLysGlnGlyTyr 329
Db 2958 ATCCAGAGCTTCGGATGTCGGGCCATTTGACTATGACCCACTTGGACAGCAACTCTAT 3017
Qy 330 GlnLeuArgGlnGlyAspAsnPheGluValTrpGluArgProLeuSerGlyLeuAlaTrp 349
Db 3018 TGGATTGACTCAGCACAAAC----- 3038
Qy 350 AlaValAlaMetIleAsnArgGlnGluIleGlyProArgSerTyrThrIleAlaVal 369
Db 3039 -----ATGATCCGAAGGACACAGAAGATGGCAGCCAGGCTTTACTGTGGTTGT 3089
Qy 370 AlaSerLeuGlyLysGlyValAlaCysAsnProAlaCysPheIleThrGlnLeuLeuPro 389
Db 3090 AGCTCAGTT-----CCGAGTCAGAACCTG---GAATACAAACC 3125
Qy 390 ValLysArgLysLeuGlyPheTyrGlu-----TrpThrSerArgLeuArgSer 405
Db 3126 TATGACCTCAGCATTTGATATTATACAGCGCTACATCTACTGAGCTTGTGAGGCTACCAAT 3185
Qy 406 HisIleAsnProThr-----GlyThrValLeu 414
Db 3186 GTCAATTAATGTCAAGATTAGATGGGAGATCAGTTGGATGTGTGTG 3233
```

RESULT 15

```
US-10-750-185-56545/c
; Sequence 56545, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KEER, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
```

```
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 56545
; LENGTH: 1615
; TYPE: DNA
; ORGANISM: Bovine 19866880768439
US-10-750-185-56545
```

Alignment Scores:

```
Pred. No.: 16.1 Length: 1615
Score: 86.50 Matches: 94
Percent Similarity: 31.52% Conservatives: 51
Best Local Similarity: 20.43% Mismatches: 173
Query Match: 3.77% Indels: 144
DB: 6 Gaps: 21
```

US-10-602-220-16 (1-417) x US-10-750-185-56545 (1-1615)

```
Qy 23 SerTrpAspIleProGlyAlaArgAlaLeuAsnGly----- 35
Db 1435 TCCTGGTAGCTCCTGGGCTGATTCCACAGGCGAGTGGCTCTGTGGTCGGGTGGAATTG 1376
Qy 36 -----LeuAlaArgThrProThrMet 42
Db 1375 CAGGAAATGAGCAAAAAATTTCCCAAGTCAGGAGGAAAGCAGAGTTCCCAACTCTC 1316
Qy 43 GlyTrpLeuHisTrp---GluArgPheMetCysAsnLeuAspCysGlnGluGlu----- 59
Db 1315 AGCCACAGAGGTGGCATGAGAGTTGCTACTCTCTGAAATTTGTTGGAAGCAGGGTTCG 1256
Qy 60 -----ProAspSerCysIleSerGluLysLeuPheMetGlu 71
Db 1255 CCATCTCTCTCCGGGACCTGAGCAATGTCATCATCAAGGGGGCTTTTATGCAG 1196
Qy 72 MetAlaGluLeuMetValSerGluGlyTrpLysAspAlaGlyTyrGluTyrLeuCysIle 91
Db 1195 AGATTTCATGAGGCCCTAGGGGAAAAATTTGGCAATTTCTCT-----TAC 1154
Qy 92 AspAspCysTrpMetAlaProGlnArgAspSerGluGlyArgLeuGlnAlaAspPro--- 110
Db 1153 GATGATTATTTATAGGCGATTGGATGATGGGGATACGAGGCTAAGGTGCTCCATT 1094
Qy 111 GlnArgPheProHisGlyIleArgGlnLeuAlaAsnTyrValHisSerLys----- 127
Db 1093 GTGAGAAGCCACACAGGAGGAGCTTGTCTGCTTTCGA-CATTTCAAGAGCGGCGACTC 1035
Qy 127 ----- 127
Db 1034 CTACAGGAGCTGCTTTTCTTTTAGAAGTTTACCTTTAATGTTCTATTATGCGCT 975
Qy 128 -----GlyLeuLysLeuGlyIleTyr 134
Db 974 TGTCTGAGCAGGCTGGGAGGAGTAGAGTAGCGGTGCGGGGGCTTTTGGGGAGAGACAG 915
Qy 135 AlaAspValGlyAsnLysThrCysAla---GlyPheProGlySerPheGlyTyrTyrAsp 153
Db 914 GCGGGGTGGTCACTCACTGAGTGGCCAGCTTGGGGCGGCGGCGGAGCTGGCAGCGGGCG 855
Qy 154 IleAspAlaGlnThrPheAlaAspTrpGlyValAspLeuLeuLysPheAspGlyCysTyr 173
Db 854 AGAGCTGTGTCAAG-----GACTGG----- 834
Qy 174 CysAspSerLeuGlnAsnLeuAlaAspGlyTyrLysHisMet---SerLeuAlaLeuAsn 192
Db 833 -----CAGCCCGCGGCCAGCAGCGCCAGCCACCTCCGCGGCTGAGCTGCAGG 786
Qy 193 ArgThrGlyArgSerIleValTyrSerCysGluTrpProLeuTyrMetTrpProPheGln 212
Db 785 CGGCCGGGCGACGAGCT-GTTTACTCTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 733
Qy 213 LysProAsnTyrThrGluLeuArgGlnTyrCysAsnHisTrpArgAsnPheAlaAspIle 232
```

Db 732 ---CCCGGCTCCACCGGCGAGCAGCAGCCTTTTCAACAATCCTGG----- 694
QY 233 AspAspSerTrpLysSerIleLysSerIleLeuAspTrpThrSerPheAsnGlnGluArg 252
Db 693 -----CTCTGGCCTGGCCCAAGGCCCTTTTCTCACTATGGACTCTGGGGCTGTGGACAGA 640
QY 253 IleValAspValAlaGlyProGlyGlyTrp-----AsnAspProAspMetLe 268
Db 639 ATCCCAAACTTGAAG---CCCCGGGCTGGATTGTCTCCGCTCTCTGTGACTGACTCACT 583
QY 268 uValIleGlyAsnPheGlyLeuSerTrpAsnGln---GlnValThrGlnMetAlaLeuTr 287
Db 582 AGGACCCCGGAGCGCTGGGCTCAGTTTCCCATGTGTCAATAATCAGATTGATGATTTG 523
QY 287 palatle-MetAlaAlaProLeuPheMetSerAsnAspLeuArgHisIleSerProGlnA 307
Db 522 GCAATGTGTTGGCAGTAGGAGAGTTTCTTCAATGATGTCCAGTGTCCAGGCCCAGAAAT 463
QY 307 lalysAlaLeuLeuGlnAspLysAspValIleAlaIleAsnGlnAspProLeuGlyLysG 327
Db 462 ACAAGCA-----GATAGCTCAGGTGGCC 439
QY 327 lnglyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrpGluArgProLeuSer---G 346
Db 438 CTGGTCACACCTGGCTCAAGGG-----CCTCTCTCTCTGGCCCTTAC 397
QY 346 lylLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGlyGly-----ProA 363
Db 396 TGCTGCCCTGGAAGCTCCTC-----CAGTTTGGGGCTGGGGAGTCCCT 352
QY 363 rgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnProAlaCysP 383
Db 351 GTGGTCCCCCCTAGCAAGCAGTCAGGCAGGAGGAGCCACTGGCTGTGTGTGTGTGTG 292
QY 383 heIleThrGlnLeuProValLysArgLysLeuGlyPheTyrGluTrpThrSer 401
Db 291 TGCTGACCCAGAGGACACACACTGCCCCAGGCTGGGT-----TGGAGCTCG 245

Search completed: January 1, 2006, 09:55:33
Job time : 1264.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: January 1, 2006, 03:35:43 ; Search time 1492.57 Seconds
(without alignments)
2310.327 Million cell updates/sec

Title: US-10-602-220-16
Perfect score: 2293
Sequence: 1 MQLRNFELHGLCALRFLA.....EWTSLRSHINPTGTVLQL 417

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US10602220/runat_30122005_140601_5538/app_query.fasta_1.782
-DB=Published Applications NA Main -QPM=fastap -SUFFIX=p2n.rnpbm
-MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1
-MATRIX=blosum62 -TRANS=human40.cd1 -LIST=45 -MODE=LOCAL -OUTFMT=ptc -NORM=ext
-THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext
-HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10602220@cgn 1.1 1364 @runat_30122005_140601_5538 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA Main:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
3: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2293	100.0	1254	3	US-09-993-059-15
2	2293	100.0	1254	5	US-10-103-327-15
3	2293	100.0	1254	7	US-10-602-219-15
4	2293	100.0	1254	7	US-10-602-220-15
5	2293	100.0	1254	8	US-10-851-388-15
6	2293	100.0	1254	3	US-10-984-389-15
7	2293	100.0	1266	3	US-09-993-059-11

ALIGNMENTS

RESULT 1
US-09-993-059-15
; Sequence 15, Application US/09993059
; Publication No. US2002008024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; TITLE OF INVENTION: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-09-993-059-15

Alignment Scores:
Pred. No.: 3,73e-275 Length: 1254
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

8	2293	100.0	1266	5	US-10-103-327-11	Sequence 11, Appl
9	2293	100.0	1266	7	US-10-602-219-11	Sequence 11, Appl
10	2293	100.0	1266	7	US-10-602-220-11	Sequence 11, Appl
11	2293	100.0	1266	8	US-10-851-388-11	Sequence 11, Appl
12	2293	100.0	1266	9	US-10-984-389-11	Sequence 11, Appl
13	2293	100.0	1272	3	US-09-993-059-17	Sequence 17, Appl
14	2293	100.0	1272	5	US-10-103-327-17	Sequence 17, Appl
15	2293	100.0	1272	7	US-10-602-219-17	Sequence 17, Appl
16	2293	100.0	1272	7	US-10-602-220-17	Sequence 17, Appl
17	2293	100.0	1272	8	US-10-851-388-17	Sequence 17, Appl
18	2293	100.0	1272	3	US-10-984-389-17	Sequence 17, Appl
19	2293	100.0	1278	9	US-09-993-059-7	Sequence 7, Appl
20	2293	100.0	1278	5	US-10-103-327-7	Sequence 7, Appl
21	2293	100.0	1278	7	US-10-602-219-7	Sequence 7, Appl
22	2293	100.0	1278	7	US-10-602-220-7	Sequence 7, Appl
23	2293	100.0	1278	8	US-10-851-388-7	Sequence 7, Appl
24	2293	100.0	1278	9	US-10-984-389-7	Sequence 13, Appl
25	2293	100.0	1284	3	US-09-993-059-13	Sequence 13, Appl
26	2293	100.0	1284	5	US-10-103-327-13	Sequence 13, Appl
27	2293	100.0	1284	7	US-10-602-219-13	Sequence 13, Appl
28	2293	100.0	1284	7	US-10-602-220-13	Sequence 13, Appl
29	2293	100.0	1284	8	US-10-851-388-13	Sequence 13, Appl
30	2293	100.0	1284	9	US-10-984-389-13	Sequence 13, Appl
31	2293	100.0	1290	3	US-09-993-059-3	Sequence 3, Appl
32	2293	100.0	1290	5	US-10-103-327-3	Sequence 3, Appl
33	2293	100.0	1290	7	US-10-602-219-3	Sequence 3, Appl
34	2293	100.0	1290	7	US-10-602-220-3	Sequence 3, Appl
35	2293	100.0	1290	7	US-10-411-037-67	Sequence 67, Appl
36	2293	100.0	1290	7	US-10-411-026-67	Sequence 67, Appl
37	2293	100.0	1290	7	US-10-410-962-67	Sequence 67, Appl
38	2293	100.0	1290	7	US-10-410-949-67	Sequence 67, Appl
39	2293	100.0	1290	7	US-10-410-930-67	Sequence 67, Appl
40	2293	100.0	1290	7	US-10-410-997-67	Sequence 67, Appl
41	2293	100.0	1290	7	US-10-411-012-67	Sequence 67, Appl
42	2293	100.0	1290	7	US-10-410-913-67	Sequence 67, Appl
43	2293	100.0	1290	8	US-10-851-388-3	Sequence 3, Appl
44	2293	100.0	1290	8	US-10-410-980-67	Sequence 67, Appl
45	2293	100.0	1290	9	US-10-410-897-67	Sequence 67, Appl

US-10-602-220-16 (1-417) x US-09-993-059-15 (1-1254)	
QY	1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db	1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTGGCC 60
QY	21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db	61 CTGCTTCTCGGACATCCTCGGGCTAGAGCACTGGACAATGGATTGGCAAGGCGCT 120
QY	41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db	121 ACCATGGCTGGCTGGACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
QY	61 AspSerCysIleSerGluLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db	181 GATTCTTCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY	81 TrpIysAspAlaGlyTyTrpGluTyLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db	241 TGGAGAGATGCAGGTATAGTACCTCTGCATTGATGATGATGATGATGATGATGATGAT 300
QY	101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db	301 GATTTCAGAGGCGAGCTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
US-10-103-327-15	
; Sequence 15, Application US/10103327	
; Publication No. US20030106095A1	
; GENERAL INFORMATION:	
; APPLICANT: GARGER, Stephen A.	
; APPLICANT: TURPEN, Thomas H.	
; APPLICANT: KUMAGAI, Monto H.	
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN	
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION	
; FILE REFERENCE: 008010087CPUS06	
; CURRENT APPLICATION NUMBER: US/10/103,327	
; CURRENT FILING DATE: 2002-03-20	
; PRIOR APPLICATION NUMBER: US/09/993,059	
; PRIOR FILING DATE: 2001-11-13	
; NUMBER OF SEQ ID NOS: 37	
; SOFTWARE: FastSeq for Windows Version 4.0	
; SEQ ID NO 15	
; LENGTH: 1254	
; TYPE: DNA	
; ORGANISM: Homo sapiens	
; FEATURE:	
; NAME/KEY: CDS	
; LOCATION: (1)...(1254)	
US-10-103-327-15	
Alignment Scores:	
Pred. No.:	3,73e-275
Score:	2293.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	5
US-10-602-220-16 (1-417) x US-10-103-327-15 (1-1254)	
QY	1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db	1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTGGCC 60
QY	21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db	61 CTGCTTCTCGGACATCCTCGGGCTAGAGCACTGGACAATGGATTGGCAAGGCGCT 120
QY	41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db	121 ACCATGGCTGGCTGGACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
QY	61 AspSerCysIleSerGluLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db	181 GATTCTTCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY	81 TrpIysAspAlaGlyTyTrpGluTyLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db	241 TGGAGAGATGCAGGTATAGTACCTCTGCATTGATGATGATGATGATGATGATGATGAT 300
QY	101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db	301 GATTTCAGAGGCGAGCTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

```
QY 121 AlaAsnTyrValHisSerIysGlyLeuIysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAAATATGTTACAGCAAGGACTGAAGCTAGGATTTATGTCAGATGTTGGAAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCCAGGCTTCCCTGGAGTTTGGATACGACATTTGATGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGACTGCTAAAATTTGATGTTGTTACTGTGACAGTTTGGAAAATTG 540
QY 181 AlaAspGlyTyrIysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAAGCACATGCTCCCTGGCCCTGAAATAGGACTGGCAGAGCATTTGTGAC 600
QY 201 SerCysGlnTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAGCCCAATATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATCTCTGGAAAAGTATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAACACGAGAGAAATTTGATGTTGTGGACAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db 781 GGTGTGAATGACCCAGATATGTTAGTATGTTGGAACTTTGGCCCTCAGCTGGAATCAGAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGCTATCATGCTGCTCTTTATTCATGTTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTTCAAGGATAAGGACGTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGACCCCTTGGCAGCAAGGGTACCAGCTTAGACGGGAGACAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATATAACCGGACGAGATGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGTAAAGAGTGGCTGTAAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValIysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCCCTGTGAAAGGAAGCTAGGGTCTATGATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTA 1251

RESULT 3
US-10-602-219-15
; Sequence 15, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: USBC-0087-CF09B
```

```
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-219-15
```

```
Alignment Scores:
Pred. No.: 3,73e-275 Length: 1254
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0
```

US-10-602-220-16 (1-417) x US-10-602-219-15 (1-1254)

```
QY 1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATCCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTCTCGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTTCATCATGAGAGAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGCAGGTTATGATACCTCTGCATTTGATGACTGTTGGATGGCTCCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAAAGGACAGACTTCAGGCAGACCCCTCAGCGCTTCTCTCATCGGATTCGCGACGTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTCCACAGCAAGGACTGAAGCTAGGGAATTTATGCAGATGTTGAAATPAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCCTGGAGTTTGGATACTACGACATTCATGCCAGACCTTTGCT 480
```

```
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
|
|
|
Db 481 GACTGGGAGTAGACTCTGTAATAATTTGATGGTTGTACTGTGACAGCTTTGGAAAATTTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
|
|
|
Db 541 CGAGATGGTTAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATTTGTATC 600

QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeuArg 220
|
|
|
Db 601 TCTGTGAGTGGCTCTTATATGTGGCCCTTCAAAAGCCCAATTTATACAGAAATCCGA 660

QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
|
|
|
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGCTGCATTTGATGATTTCTCTGGAAAAGTAAAG 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
|
|
|
Db 721 AGTATCTTGGACTGGGACATCTTTTAAACAGGAGAGAAATTTGTATGTTCTGGACCAGG 780

QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
|
|
|
Db 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840

QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
|
|
|
Db 841 GTAACCTCAGATGGCCCTCTGGGGTATCATGGCTGCTCCTTTATTCATGTCTAATGACCTC 900

QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
|
|
|
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAT 960

QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlnGlyAspAsnPheGluValTrp 340
|
|
|
Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGGACAACTTTGAAGTGTGG 1020

QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 360
|
|
|
Db 1021 GAACGACCTCTCTCAGGCTTAGCGCTGGCTGTAGCTATGATAAACCCGCGAGAGATGGT 1080

QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
|
|
|
Db 1081 GGACCTCGCTCTTATACCATCGGAGTTGCTTCCTCGGTAAAGAGTGGCTGTATCTCT 1140

QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
|
|
|
Db 1141 GCCTGCTTCATCACAGACTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200

QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
|
|
|
Db 1201 TCAAGGTTAAGAAATCACAATAATCCACAGGCACTGTTTTGTCTCAGCTA 1251

RESULT 4
US-10-602-220-15
; Sequence 15, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP07B
; CURRENT APPLICATION NUMBER: US/10/602,220
; PRIOR FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
```

```
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-220-15

Alignment Scores:
Pred. No.: 3,73e-275 Length: 1254
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 7

US-10-602-220-16 (1-417) x US-10-602-220-15 (1-1254)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
|
|
|
Db 1 ATGCAGCTCAGGAACCCAGAACTACATCTGGGCTGCGCGCTGCGCTTGCCTTCCCTGGGCC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
|
|
|
Db 61 CTCGTTTCTCTGGGACATCTCTGGGGCTAGAGCACTGGACATGTCATGTCAGTCAGGAGCGCCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
|
|
|
Db 121 ACCATGGCTGGCTGCTGCTGAGGAGCGCTTTCATGTGCAACCTTGCATGCCAGGAGAGCCA 180

QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
|
|
|
Db 181 GATTCTCTCATCATGTCAGTGAAGACTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGGC 240

QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
|
|
|
Db 241 TGGAGAGATGCAGGTATGATGATACCTTCATTTGATGACTGTTGGATGGCTCCCAAGA 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
|
|
|
Db 301 GATTTCAGAGGCGAGACTTTCAGGCGAGCCCTCAGCGCTTTCCTCATGGGATTCGCGACGCTA 360

QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
|
|
|
Db 361 GCTAATTTATGTTTACAGCAAGGAGCTGAACTAGGGATTTATGCAGATGTTGGAAATATA 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
|
|
|
Db 421 ACCTGCCAGGCTTCCCTCGGAGATTTTGGATACACTACGACATTCATGCCAGACCTTTGCT 480

QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
|
|
|
Db 481 GACTGGGAGTAGACTCTGCTAAATTTGATGGTTGTACTGTGACAGCTTTGGAAAATTTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
|
|
|
Db 541 GCAGATGGTTATAAGCACATGTCTCTGGCCCTGAATAGGACTGGCAGAAAGCTTTGTGTAC 600
```

```
QY 201 SerCysGluTrpProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluLeuArg 220
Db 601 TCCTGTGAGTGGCCCTCTTATATATGTGGCCCTTTCAAAAGGCCCAATTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGCGGAAATTTTGTGACATGATGATTCCTGGAAAGATATTAAG 720
QY 241 SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTATGTGTGGACCAAGG 780
QY 261 GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTyrAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db 961 CAGGACCCCTTGGCAAGCAAGGCTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATATAACCGGAGGAGATGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTATACCATCGCAGTGTCTTCCCTGGTAAAGAGTGGCTGTATCTCT 1140
QY 381 AlaCysPheIleThrGlnLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db 1141 GCCTGTCTCATCACAGCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTAAAGAGTCACATAAATCCACAGGCACTGTTTTGTCTTCAGCTA 1251
```

RESULT 5

```
; Sequence 15, Application US/10851388
; Publication No. US20040234516A1
; GENERAL INFORMATION:
; APPLICANT: KUMAGAI, Tomohiko H.
; APPLICANT: TURPEN, Stephen A.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851.388
; CURRENT FILING DATE: 2004-05-21
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-10-851-388-15
```

```
Alignment Scores: 3.73e-275 Length: 1254
Pred. No.: 2293.00 Matches: 417
Score:
```

```
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0
US-10-602-220-16 (1-417) x US-10-851-388-15 (1-1254)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCGCTTCCTGGCC 60
QY 21 LeuValSerTyrAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTCTCTGGGACATCTCCCTGGGCTTAGAGCACTGGACAATGGATGGCAAGGACCCCT 120
QY 41 ThrMetGlyTyrLeuHisTyrGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCACCTTGTACTGCCAGGAGAGACCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGGATGCAAGTTATGAGTACCTCTGCATGTAGTACTGTTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAGGAGAGACTTCAGGAGAGCCCTCAGCGCTTCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTTATGTTTACAGCAAGGAGCTGAAGCTAGGCAATTTATGCAGATGTTGGAATAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATATCTACGACATTTGATGTCAGACATTTGGCT 480
QY 161 AspTyrGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGTAATAATTTGATGGTGTGTACTGTGACAGCTTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTCAATAGGACTGGCAGAAATTTGTGTAC 600
QY 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCCCTCTTTATATATGTGGCCCTTTCAAAAGGCCCAATTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGCGGAAATTTTGTGACATGATGATTCCTGGAAAGATATTAAG 720
QY 241 SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTATGTGTGGACCAAGG 780
QY 261 GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTyrAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db 961 GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGGATGCAAGTTATGAGTACCTCTGCATGTAGTACTGTTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAGGAGAGACTTCAGGAGAGCCCTCAGCGCTTCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTTATGTTTACAGCAAGGAGCTGAAGCTAGGCAATTTATGCAGATGTTGGAATAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATATCTACGACATTTGATGTCAGACATTTGGCT 480
QY 161 AspTyrGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGTAATAATTTGATGGTGTGTACTGTGACAGCTTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTCAATAGGACTGGCAGAAATTTGTGTAC 600
QY 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCCCTCTTTATATATGTGGCCCTTTCAAAAGGCCCAATTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGCGGAAATTTTGTGACATGATGATTCCTGGAAAGATATTAAG 720
QY 241 SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTATGTGTGGACCAAGG 780
QY 261 GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTyrAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db 961 GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGGC 240
```

Db 961 CAGGACCCCTTGGCAGCAAGGTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCGCGAGGAGATTGGT 1080
Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCTCGGGTAAAGAGTGGCCTGTAAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValIleArgLysLeuGlyPheTrpGluTrpThr 400
Db 1141 GCCTGCTTCAACACAGCTCCTCCTGTGAAAGGAGTAGGTTCTATGAATGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTA 1251

RESULT 6

US-10-984-389-15

; Sequence 15, Application US/10984389

; Publication No. US20050125859A1

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; APPLICANT: KUMAGAI, Monto H.

; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN

; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION

; FILE REFERENCE: 008010087CPUS06

; CURRENT APPLICATION NUMBER: US/10/984,389

; CURRENT FILING DATE: 2004-11-08

; PRIOR APPLICATION NUMBER: US/09/993,059

; PRIOR FILING DATE: 2001-11-13

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: Fast-Seq for Windows Version 4.0

; SEQ ID NO 15

; LENGTH: 1254

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(1254)

US-10-984-389-15

Alignment Scores:

Pred. No.:	3,73e-275	Length:	1254
Score:	2293.00	Matches:	417
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-10-602-220-16 (1-417) x US-10-984-389-15 (1-1254)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTCGGCTGCGCGCTTCGCTTCCTCGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyValArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTGCATCAGTGAAGAGCTCTTCATGGAGATGCCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGNAGGATCGAGGTATAGTACCTCTGCATTGATGACTGTGGATGGCTCCCCAAGA 300

RESULT 7

US-09-993-059-11

; Sequence 11, Application US/09993059

; Publication No. US20020088024A1

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; APPLICANT: KUMAGAI, Monto H.

Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAAAGGAGCACTTCAGCAGACCTCTAGCGCTTCTCATGGGATTGCCAGCTA 360
Qy 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTTCACAGCAAGGACTGAAGCTAGGATTTATGCAGATGTTGGAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCTCGGAGTTTGGATATACACATTCATGATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGCTTTGGAAATTTG 540
Qy 181 AlaAspGlyTyzLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
Db 541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTGATAGGACTGGCAGAAAGCATTTGTATC 600
Qy 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
Db 601 TCCTGTAGTGGCCTCTTTATATATGTCGCCCTTTCAAAGGCCCAATTTATACAGAAATCCGA 660
Qy 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACATGGCGAAATTTTGTCTGACATTCATGATGATTCCTGGAAAGATTAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAACACAGGAGAGAAATTTGTATGCTGTCGACCAAGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAATGACCCAGATATGTTAGTGATGGCACTTTGGCCCTCAGCTCGGATAGCAATCAGCAA 840
Qy 281 ValTrpGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGCTTAATGACCIC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGATAAGGACGCTAATTTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACCAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATATAAACCGCGAGGAGATTGGT 1080
Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTGGTAAAGAGTGGCTCTGTAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
Db 1141 GCCTGCTTCAACACAGCTCCTCCCTGTGAAAGGAAAGCTAGGGTTCTATGAATGAGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTA 1251

;; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
;; FILE REFERENCE: 008010087CPUS06
;; CURRENT APPLICATION NUMBER: US/09/993,059
;; CURRENT FILING DATE: 2001-11-13

;; NUMBER OF SEQ ID NOS: 37
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 11
;; LENGTH: 1266

;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(1266)
US-09-993-059-11

Alignment Scores:

Pred. No.: 3,78e-275 Length: 1266
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 0 Gaps: 3

US-10-602-220-16 (1-417) x US-09-993-059-11 (1-1266)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCGGACATCCTCGGGCTAGAGCACTGGCAATGATGGCAAGAGCGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGGACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTCAGGAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGAAGGATGCAGGTTATGAGTACCTCTGCAITGATGACTGTGGATGGCTCCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAAGGCAGACTTCAGGAGACCTTCAGCGCTTTCCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTyxAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGAGACTGAAGCTAGGATTTATGCAGATGTTGGAAATAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTyxAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCTCGGAGTTTGGATATACACATTCATGATGATGATGATGATGAT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyxAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTATGATCTGCTAAATTTGATGGTTGTTACTGTGACAGATTTGGAAATTTG 540
Qy 181 AlaAspGlyTyxLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyx 200
Db 541 GCAGATGGTTATAGCACATGCTTCCTGGCCCTGAATAGACTGGCAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyxMetTrpProPheGlnLysProAsnTrpThrGluLeuArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAAGCCCAATATATACAGAAATCCGA 660
Qy 221 GlnTyxCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCAGTCGGCGAAATTTTGTGATGATGATGATGATGATGATGATGATGATG 720

Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGAGCTGGACATCTTTTAAACAGGAGAGAATTTGTGATGTTGCTGACAGGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db 781 GGTGGAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTATGTCATATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGCTAATTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAACCGGACGAGATTGGT 1080
Qy 361 GlyProArgSerTyxThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCTGGGTAAAGGAGTGCGCTGTAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyxGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAAGGAAGCTAGGGTTCTATGATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAGGTTAAGAGTCACATAAATCCACAGCGACTGTTTTTGTCTCAGCTA 1251

RESULT 8

US-10-103-327-11
; Sequence 11, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-10-103-327-11

Alignment Scores:

Pred. No.: 3,78e-275 Length: 1266
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 0 Gaps: 5

US-10-602-220-16 (1-417) x US-10-103-327-11 (1-1266)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlnLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTTGGCAAGGAGCCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCATCGGAGCGCTTCATGTGCACACCTTGACTGCCAGGAGGCCA 180
QY 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTGAGAAGCTTCTCATGGAGATGGCAGAGCTCATGTGCTCAGAAAGC 240
QY 81 TrpLysAspAlaGlyTyrgluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGAAGGATGCAGGTTATGAGTACCTCTGCATTTGATGACTGTGGATGGCTCCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAAGAGCAGACTTCAGCAGACCTCAGCGCTTCTCTCATGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTTATGTTCCACAGCAAGGACTGAAGCTAGGGAATTTATGCAGATGTGGAAATAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGANATCTAGCAGATTCATGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTACAGATTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATTAAGCACATGTCTTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAGGCCAATATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCATCTACTGGCGAATTTTGTCTGACATTTGATGATCTTCTGGAAAGATATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGACTGGACATCTTTAAACCAGGAGAGATTTGTATGTTGTGGACCAAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrglnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGACCCCTTGGGCAAGCAAGGGTACCGAGCTTAGACAGGAGACAACCTTTGAATGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnIleGly 360
Db 1021 GAAACGACCTCTCTCAGGCTTAGCCCTGAGCTATGATATAAACCCGCGAGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380

Db 1081 GGACCTCGCTCTTATACCATCGAGTCTTCCCTGGGTAAAGGAGTGGCCTGTAACTCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCTCTCCCTGTGTAAGAAAGAGCTAGGCTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTACATAAATCCACAGGCACTGTTTGTCTTCAGCTA 1251

RESULT 9
US-10-602-219-11
; Sequence 11, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-219-11

Alignment Scores:
Pred. No.: 3 78e-275 Length: 1266
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-10-602-220-16 (1-417) x US-10-602-219-11 (1-1266)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlnLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTTGGCAAGGAGCCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60

Db 121 ACCATGGCTGGCTGCATCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCTGCATCAGTGAAGAGCTCTTCTGAGAGTGGCAGAGCTCATGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGAGGTTATGAGTACTCTGCATTGATGACTTGTGGTGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTGAGAGGAGACTTCAGCAGACCTCAGCGCTTCTCATGGATTGCGCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGAGCTGAAGCTAGGATTTATGAGATTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCCAGGCTTCCCTGGAGATTTGGATACCTACGACATTTGATGCCAGACCTTGGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGACAGCTTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGTTATAGACCATGCTCTTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeuArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGGCCAATTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTCGGAAATTTGTGACATTTGATGATTCCTCGAAAGATATAAG 720
Qy 241 SerIleLeuAspTrpTrpSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTGATTTGTGGACACAGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGCTTATCATGGCTGCTCTTTATTCATGTTCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCCAAGCTCTCTTCAGGATTAAGACGCTTAATGGCAATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGACCCCTTGGCAAGCAAGGTACCAAGCTAGACAGGAGACCACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTACCTGGCTGTAGCTATGATATAAACCCGCGAGGAGATTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCTCTGGGTAAAGAGTGGCTCTGTAATCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGATGAGCT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417

Db 1201 TCAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTGTCTTACAGCTA 1251
RESULT 10
US-10-602-220-11
; Sequence 11, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Poque, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: USBC-0087-CP07B
; CURRENT APPLICATION NUMBER: US/10/602,220
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-220-11
Alignment Scores:
Pred. No.: 3,78e-275 Length: 1266
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0
US-10-602-220-16 (1-417) x US-10-602-220-11 (1-1266)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGACCCAGACACTACATCTGGGCTGGCGCTTGGCTTCTGGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGACATCTCCCTGGGCTAGAGCACTGCACATGATGATGGCAAGACGCCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluPro 60
Db 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCTGCATCAGTGAAGAGCTTCTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100

```
Db 241 TCGAAGGATGCGAGGTATCAGTACCTCTCATGTGATGCTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAAGAGGCAGACTTCAGGCAGACCTTCAGCGCTTCTCCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAAATATGTTACACAGCAAGGACTGAAGCTAGGGAATTAATGCAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGGCTTCCTCGGAGTTTGGATACACTACGACATGATGATGCCAGACCTTGTCT 480
Qy 161 AspTyrGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTCGAAAAATTGG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTCAATAGGACTGGCAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCCTCTTTATATATGTGGCCCTTTCAAAAGGCCCAATTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
Db 661 CAGTACTGCATCCTCGGCGAATTTTGTGACATTTGATGATTCCTGGAAAGATATAAG 720
Qy 241 SerIleLeuAspTyrThrSerPheAsnGlnLysArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACCCAGG 780
Qy 261 GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGlnGln 280
Db 781 GGTGTGAATGACCCAGATATGTTAGTGATTTGGCAACTTTGGCCTCAGCTGGAAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTyrAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCATAAGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCGCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACAGCTTAGCAGGAGCAAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAAGCACTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCCGCGAGGATTTGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCCCTGTAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGTGAAGAGGAGCTAGGGGTCTCATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGTTTAAGAGTCATATAATCCCAAGGCACTGTTTGTCTTCAGCTA 1251
```

RESULT 11

US-10-851-388-11

; Sequence 11, Application US/10851388

; Publication No. US20040234516A1

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

```
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851,388
; CURRENT FILING DATE: 2004-05-21
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1) ... (1266)
US-10-851-388-11
```

```
Alignment Scores:
Pred. No.: 3,78e-275 Length: 1266
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
```

US-10-602-220-16 (1-417) x US-10-851-388-11 (1-1266)

```
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGACCCAGACACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTCTGGCC 60
Qy 21 LeuValSerTyrAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGGTTCTCTGGACATCCCTGGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTyrLeuHisTyrGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCACCTTGACTGCCAGGAAGAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTyrPheMetAlaProGlnArg 100
Db 241 TGGAAAGGATGCAGGTTATGAGTACCTCTGCATTGATGACTGTGTGGATGGCTCCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAGGCGAGACTTCAGGCAGACCCCTCAGGGCTTTCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAAATATGTTACACAGCAAGGACTGAAGCTAGGGAATTAATGCAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGGCTTCCTCGGAGTTTGGATACACTACGACATGATGATGCCAGACCTTGTCT 480
Qy 161 AspTyrGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTCGAAAAATTGG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTCAATAGGACTGGCAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCCTCTTTATATATGTGGCCCTTTCAAAAGGCCCAATTATACAGAAATCCGA 660
```

```
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspSerTrpLysSerIleLys 240
|||
DB 661 CAGTACTGCAATCACTGCGCAATTTTGGCTGACATTGATCTCTGGAAAAGTATAAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnArgIleValAspValAlaGlyProGly 260
|||
DB 721 AGTATCTTGACCTGGACATCTTTTAAACGAGAGAAATTGTGATGTCTGGACCAAGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
|||
DB 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
|||
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
|||
DB 901 CGACATCATGACCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATGGCCATCAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
|||
DB 961 CAGGACCCCTTGGCAAGCAAGGTACCACTTAGACAGGAGACAACCTTTGAAGTGTG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
|||
DB 1021 GAACGACCTCTCTCAGGCTTAGCGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080
QY 361 GlyProArgSerTrpThrIleAlaValAlaSerIleGlyLysGlyValAlaCysAsnPro 380
|||
DB 1081 GGAACCTCGCTCTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
|||
DB 1141 GCCTGCTTCATCACAGACTCTCTCCCTGTGAAAGGAGCTAGGGTTCATGATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
|||
DB 1201 TCAAGGTGAAGAGTCACATAAATCCCAAGGCACTGTTTGTCTTCAGCTA 1251
```

RESULT 12

```
US-10-984-389-11
; Sequence 11, Application US/10984389
; Publication No. US20050125859A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/984,389
; CURRENT FILING DATE: 2004-11-08
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-10-984-389-11
```

Alignment Scores:

Pred. No.:	3,78e-275	Length:	1266
Score:	2293.00	Matches:	417
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0

```
DB: 9 Gaps: 0
US-10-602-220-16 (1-417) x US-10-984-389-11 (1-1266)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
|||
DB 1 ATCAGCTGAGGAACCCAGAACCTACATCTGGGCTGCGGCTTGGCTTCGCTTCTGCGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
|||
DB 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATGGCAAGGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
|||
DB 121 ACCATGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
|||
DB 181 GATTCTTCATCAGTGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
|||
DB 241 TGGAAGGATGCAGGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
|||
DB 301 GATTCAAGAGGACAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
|||
DB 361 GCTAATTATGTTCAACAGCAAGGACTGAGCTAGGAGATTTATGCAGATGTTGGAAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
|||
DB 421 ACCTGCGCAGGCTTCTCTGGAGTTTGGATACCTACGACATTGATGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
|||
DB 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGCACAGTTTGGAAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
|||
DB 541 GCAGATGGTTATGAACACATGCTCTTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
|||
DB 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAAGCCCAATTATATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
|||
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCACATTTGATGATTCCTGGAAAAGTATAAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
|||
DB 721 AGTATCTTGACCTGGACATCTTTTAAACGAGAGAAATTGTGATGTCTGGACCAAGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
|||
DB 781 GGTGGAAATGACCCAGATATGTTAGTGAATTTGGCAACTTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
|||
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
|||
DB 901 CGACATCATGACCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATGGCCATCAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
|||
DB 961 CAGGACCCCTTGGCAAGCAAGGTACCACTTAGACAGGAGACAACCTTTGAAGTGTG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
```

Db 1021 GAACGACCTCTCTCAGCGCTTAGCGCTGGCTGTAGCTATGATTAACCGGAGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGAACCTCGCTCTTATACCATCGAGTGTCTCCCTGGGTAAAGGAGTGGCTGTATCTCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db 1141 GCCTGCTTCATCACAGACTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTCAATAAATCCACAGGCACCTGTTTTGCTTCAGCTA 1251

RESULT 13

US-09-993-059-17
; Sequence 17, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1272
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1272)
US-09-993-059-17

Alignment Scores:
Pred. No.: 3,81e-275 Length: 1272
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-16 (1-417) x US-09-993-059-17 (1-1272)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCTTCGCTTCCTCGGCC 60
QY 21 LeuValSerTrpAspIleProGlyValArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCGGGACATCCCTGGGGCTAGAGCATCTGCACATGATGTTGGCAGGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCATCGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGATGAGGTATAGTACTCTGCAATGATGACTTGTGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAGGCAGACTTCAGCGAGACCCCTCAGCGCTTTCCTCATGGGATTCGCGAGCTA 360

QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTTACAGCAAGAGACTGAAGCTAGGATTTATGACAGATGTCGAAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGTGATATACGACATTTGATGTCGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGACTCTGCTAAAATTTGATGGTTGTTACTGTGACAGATTTGGAAATTTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTTATAGCACATGCTTGTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAGGCCCAATTTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACCTGGCGAAATTTGCTGCAATTTGATTCCTGGGAAAGATATAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAACAGGAGGAGAAATTTGTGATGTTGCTGACACGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAATGACCCAGATATGTTAGTGTGGCAACTTTGGCCCTCAGCTCGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCATATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACAGCTTAGACAGGAGACCAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATATAAACCCGCGAGGATTTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db 1141 GCCTGCTTCATCACAGACTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTCAATAAATCCACAGGCACCTGTTTTGCTTCAGCTA 1251

RESULT 14

US-10-103-327-17
; Sequence 17, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20

;; PRIOR APPLICATION NUMBER: US/09/993,059

;; PRIOR FILING DATE: 2001-11-13

;; NUMBER OF SEQ ID NOS: 37

;; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 17

;; LENGTH: 1272

;; TYPE: DNA

;; ORGANISM: Homo sapiens

;; FEATURE:

;; NAME/KEY: CDS

;; LOCATION: (1)...(1272)

US-10-103-327-17

Alignment Scores:

Pred. No.: 3,81e-275 Length: 1272

Score: 2293.00 Matches: 417

Percent Similarity: 100.00% Conservations: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 5 Gaps: 0

US-10-602-220-16 (1-417) x US-10-103-327-17 (1-1272)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGCGCTGGCGCTTCGCTTCCTGGCC 60

Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40

Db 61 CTGGTTCTCGGNCATCCCTGGGGCTAGAGCACTGGCAATGGATTGGCAAGACGCT 120

Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60

Db 121 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGACCA 180

Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80

Db 181 GATTCTGTCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100

Db 241 TGGNAGATCAGGTTATGATGACTCTGATGATGATGATGATGATGATGATGATGATGAT 300

Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120

Db 301 GATTCAAGAGCAGACTTCAGGCAGACCTCAGCGCTTCTCATGGATTGCCAGCTA 360

Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140

Db 361 GCTAATATGTTTACAGCAAGAGCTGAAGCTAGGAGTTTATGATGATGATGATGATGATGAT 420

Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTyrAspIleAspAlaGlnThrPheAla 160

Db 421 ACTTGGCAGCTTCCCTGGAGTTTGGATGATGATGATGATGATGATGATGATGATGAT 480

Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180

Db 481 GACTGGGAGTAGACTGCTGCTAAATTTGATGGTTGTTTACTGTGACAGTTGATGATGATGAT 540

Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200

Db 541 GCAGATGGTTATAGCAGCATGCTCTTGGCCCTGAATAGGAGCTGGCAGAGCATTTGTGTAC 600

Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220

Db 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTCAAAAGGCCCAATATATACAGAAATCCGA 660

Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240

Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGATGATGATGATGATGATGATGATGATGATGAT 720

Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260

Db 721 AGTATCTTGACTGGACATCTTTTAACGAGAGAGAAATTTGATGTTGCTGGACCAAGG 780

Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280

Db 781 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840

Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300

Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCTATGTCTAATGACCTC 900

Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320

Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATGCCATCAAT 960

Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340

Db 961 CAGGACCCCTTGGCAAGCAAGGTTACCACTTAGACAGGAGACAACTTTGAAGTGTGG 1020

Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360

Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATAAACCCGCGAGAGATTGGT 1080

Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380

Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140

Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400

Db 1141 GCCTGCTTCATCACACAGCTCCCTCCCTGTGAAAGGAGAGCTAGGTTCTATGATGAGCT 1200

Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417

Db 1201 TCAAGGTAAAGATCACATAAATCCACAGGACCTGTTTGTCTTCTAGCTA 1251

RESULT 15

US-10-602-219-17

;; Sequence 17, Application US/10602219

;; Publication No. US20040016021A1

;; GENERAL INFORMATION:

;; APPLICANT: Large Scale Biology Corporation

;; APPLICANT: Turpen, Thomas H.

;; APPLICANT: Pogue, Gregory P.

;; APPLICANT: Erwin, Robert L.

;; APPLICANT: Grill, Laurence K.

;; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION

;; FILE REFERENCE: LSBC-0087-CP09B

;; CURRENT FILING DATE: 2003-06-23

;; PRIOR APPLICATION NUMBER: 09/993,059

;; PRIOR FILING DATE: 2001-11-13

;; PRIOR APPLICATION NUMBER: 09/626,127

;; PRIOR FILING DATE: 2000-07-26

;; PRIOR APPLICATION NUMBER: 09/316,572

;; PRIOR FILING DATE: 1999-05-21

;; PRIOR APPLICATION NUMBER: 08/324,003

;; PRIOR FILING DATE: 1994-10-14

;; PRIOR APPLICATION NUMBER: 08/176,414

;; PRIOR FILING DATE: 1993-12-29

;; PRIOR APPLICATION NUMBER: 07/997,733

;; PRIOR FILING DATE: 1992-12-30

;; PRIOR APPLICATION NUMBER: 08/184,237

;; PRIOR FILING DATE: 1994-01-19

;; PRIOR APPLICATION NUMBER: 07/923,692

;; PRIOR FILING DATE: 1992-07-31

;; PRIOR APPLICATION NUMBER: 07/600,244

;; PRIOR FILING DATE: 1990-10-22

;; PRIOR APPLICATION NUMBER: 07/641,617

;; PRIOR FILING DATE: 1991-01-16

;; Remaining Prior Application data removed - See File Wrapper or PALM.

;; NUMBER OF SEQ ID NOS: 32

;; SOFTWARE: PatentIn version 3.2

;; SEQ ID NO 17

;; LENGTH: 1272

! TYPE: DNA
! ORGANISM: Homo sapiens
US-10-602-219-17

Alignment Scores:

Pred. No.: 3 81e-275 Length: 1272
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-10-602-220-16 (1-417) x US-10-602-219-17 (1-1272)

```
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTTCCTGGGACATCCCTGGGGCTAGAGCAATGGATTGGCAAGGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCTCGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGGAAAGATCAGGTTATAGTAGTACCTCTGATGTAGTACTGTTGGATGGCTCCCCAAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTTCAGAGGCAGACTTCAGGCAGACCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
DB 361 GCTAAATTATGTTACAGCAAAAGGACTGAAGCTAGGAGTTTATGCAGATGTTGGAAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGCCAGCGCTTCCCTGGGAGTTTGGATACCTACGACATTGATGCCCGCAGACTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAAATTG 540
QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
DB 541 GCAGATGGTTATAAGCAGCATGCTCCTGGCCCTGAATAGGACTGGCAGAACATTGTGTAC 600
QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
DB 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTCAAAGGCCCAATATACAGAAATCCGA 660
QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACTGGGNAATTTTGTGCTGACATTGATGATTCCTGGAAAAGTATAAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 721 AGTATCTTGGACTGGACATCTTTTAAACGAGAGAGAATTGTTGATGTTGCTGGACACGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
DB 781 GGTTCGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTCTTAATGACCTC 900
```

```
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACACATCAGCCCTCAGCCCAAGCTCTCCTTCAGATTAAGGACGTAATTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTTAGCTATATAAACCGCAGGAGATTGGT 1080
QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1081 GGACCTTCGGCTCTTATACCATCGCAGTTGCTTCCTGGGTAAAGGAGTGGCTGTATTCCT 1140
QY 381 AlaCysPheIleThrGlnLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
DB 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGMAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
DB 1201 TCAAGGTTAAGAGTCACTATAATCCACAGGCACTGTTTGTGCTTCAGCTA 1251
```

Search completed: January 1, 2006, 09:34:32

Job time : 1520.57 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2006, 00:29:37 ; Search time 339.352 Seconds
(without alignments)
2184.292 Million cell updates/sec

Title: US-10-602-220-16

Perfect score: 2293

Sequence: 1 MQLRNLHLCALALRFLA.....EWTSLRSHNPQTLLQL 417

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US10602220/runat_30122005_140559_5451/app_query.fasta_1.782
-DB=Issued Patents NA -QPMF=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=spt -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptc -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10602220@cgn 1.1.237 @runat_30122005_140559_5451 -NCPU=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.*

1: /cgn2_6/ptodata/1/ina/1 COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5 COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/H COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
7: /cgn2_6/ptodata/1/ina/PP COMB.seq.*
8: /cgn2_6/ptodata/1/ina/RE COMB.seq.*
9: /cgn2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2293	100.0	1254	3	US-09-993-059-15
2	2293	100.0	1254	3	US-10-103-327-15
3	2293	100.0	1266	3	US-09-993-059-11
4	2293	100.0	1272	3	US-10-103-327-11
5	2293	100.0	1272	3	US-09-993-059-17
6	2293	100.0	1272	3	US-10-103-327-17
7	2293	100.0	1278	3	US-09-993-059-7
8	2293	100.0	1278	3	US-10-103-327-7
9	2293	100.0	1284	3	US-09-993-059-13

10	2293	100.0	1284	3	US-10-103-327-13	Sequence 13, Appl
11	2293	100.0	1290	3	US-09-993-059-3	Sequence 3, Appl
12	2293	100.0	1290	3	US-10-103-327-3	Sequence 3, Appl
13	2293	100.0	1296	3	US-09-993-059-9	Sequence 9, Appl
14	2293	100.0	1296	3	US-10-103-327-9	Sequence 9, Appl
15	2293	100.0	1308	3	US-09-993-059-5	Sequence 5, Appl
16	2293	100.0	1308	3	US-10-103-327-5	Sequence 5, Appl
17	2293	100.0	1343	3	US-08-928-881-18	Sequence 18, Appl
18	2293	100.0	1343	3	US-09-543-921-18	Sequence 18, Appl
19	2293	100.0	1343	3	US-09-266-014-3	Sequence 3, Appl
20	2293	100.0	1343	3	US-09-491-759-18	Sequence 18, Appl
21	2293	100.0	1393	2	US-07-602-824A-1	Sequence 1, Appl
22	2293	100.0	1393	2	US-07-983-451-1	Sequence 6, Appl
23	2293	100.0	1393	2	US-08-261-577-6	Sequence 14, Appl
24	2293	100.0	8234	3	US-09-626-127-14	Sequence 34, Appl
25	2293	100.0	8234	3	US-09-993-059-34	Sequence 34, Appl
26	2293	100.0	8234	3	US-10-103-327-34	Sequence 21, Appl
27	2231	97.3	1233	3	US-09-993-059-21	Sequence 21, Appl
28	2231	97.3	1233	3	US-10-103-327-21	Sequence 19, Appl
29	2227	97.1	1215	3	US-09-993-059-19	Sequence 19, Appl
30	2227	97.1	1215	3	US-10-103-327-19	Sequence 19, Appl
31	2213	96.5	1304	9	5179023-3	Patent No. 5179023
32	2157	94.1	1214	3	US-09-023-655-1001	Sequence 1001, Ap
33	2132	93.0	11641	3	US-09-626-127-13	Sequence 13, Appl
34	2132	93.0	11641	3	US-09-993-059-33	Sequence 33, Appl
35	2132	93.0	11641	3	US-10-103-327-33	Sequence 33, Appl
36	2131	92.9	1158	3	US-09-176-666-53	Sequence 53, Appl
37	2131	92.9	1161	3	US-09-176-666-52	Sequence 52, Appl
38	2131	92.9	1164	3	US-09-176-666-51	Sequence 51, Appl
39	2131	92.9	1167	3	US-09-176-666-50	Sequence 50, Appl
40	2131	92.9	1170	3	US-09-176-666-49	Sequence 49, Appl
41	2131	92.9	1173	3	US-09-176-666-48	Sequence 48, Appl
42	2131	92.9	1176	3	US-09-176-666-47	Sequence 47, Appl
43	2131	92.9	1179	3	US-09-176-666-46	Sequence 46, Appl
44	2131	92.9	1182	3	US-09-176-666-45	Sequence 45, Appl
45	2131	92.9	1188	3	US-09-176-666-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1

US-09-993-059-15
; Sequence 15, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-09-993-059-15

Alignment Scores:
Pred. No.: 2,19e-249
Score: 2393.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 3
Length: 1254
Matches: 417
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-10-602-220-16 (1-417) x US-09-993-059-15 (1-1254)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTGGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGACATGATGGCAGGACGCTT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGCTGGCTGCACCTGGAGGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCCA 180
QY 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCTGCATCAGTCAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
QY 81 TrpLysAspAlaGlyTrpGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGAAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTCAAGAGGACACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 361 GCTAAATTATGTTCCACAGCAAGGACTGAAGCTAGGGAATTTATGCAGATGTTGGAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCAGGCTTCCTGGAGGTTTGGATACCTACGACATTGATGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGAGTAGACTCTGCTAAATTTGATGGTGTGTACTGTGACAGCTTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 GCAGATGGTTATAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
DB 601 TCTGTGAGTGGCTCTTTATATATGTGGCCCTTCAAAGGCCAATTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCTGACATTGATGATTCCTGGAAAGTATAAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 721 AGTATCTTGACTGGACATCTTTAAACGAGAGAGAAATTTGATGTGTGTCGACAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GGTGGAAATGACCCAGATATGATGATTTGGCNACTTTGGCCTCAGCTGGATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACCTCAGATGGCCCTCTGGGCTCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACACATACGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTTGGCAGCAGAGGTATCAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATATAACCGCAGGAGATTGGT 1080

QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1081 GGACCTCGCTCTTATACCATCGAGTGTCTCCCTGGGTAAAGGAGTGGCTGTATCTCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
DB 1141 GCCTGCTTCATCACAGCTCCTCCCTGTGTAAAGGAAGCTAGGGTCTTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
DB 1201 TCAAGGTTAAGAAGTACATAAATCCACAGGACTGTTTGTCTCAGCTA 1251
RESULT 2
US-10-103-327-15
; Sequence 15, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KIMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-10-103-327-15

Alignment Scores:
Pred. No.: 2,19e-249 Length: 1254
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-16 (1-417) x US-10-103-327-15 (1-1254)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTGGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGACAAATGGATTTGGCAAGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGCTGGCTGCACCTGGAGGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCCA 180
QY 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCTGCATCAGTCAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGAAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTCAAGAGGACAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140


```
Db 361 GCTAATTTATGTTACACCAAGGACTGAAGCTTAGGGATTTATGCAGATGTTGGAAATAAA 420
QY ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGCTTCCCTGGAGATTTGGATACTACGACATTTGATGCCAGACCTTTGCT 480
QY AspTyrGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAenLeu 180
Db 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTGTACTGTGCACAGTTTGGAAAATTTG 540
QY AlaAspGlyTyrLysHisMetSerLeuAlaLeuAenArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATTAAGCACATGCTCTGGCCCTGAATAGGACTGGCAGACATTTGTGTAC 600
QY SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
Db 661 CAGTACTGCAATCAGTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAAGTATAAAG 720
QY SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGACCTGGACATCTTTTAACCCAGGAGAAATTTGATGTGTGACACCGGG 780
QY GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGlnGln 280
Db 781 GGTGGNATGACCCAGATATGTAGTATTTGGCACTTTGGCTCAGCTGGAAATCAGCAA 840
QY ValThrGlnMetAlaLeuTyrAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCGGGCTATCATGGCTGCTCTTTATTTATGTCATAATGACCTC 900
QY ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
QY GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTCACGCTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATAAACCGCAGGAGATTGGT 1080
QY GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaLysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGTAAAGGAGTGGCTGTAATCCT 1140
QY AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAAGGAAGTAGGGTTCTATGATGAGACT 1200
QY SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTA 1251
```

RESULT 3

```
US-09-993-059-11
; Sequence 11, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 11
```

```
; LENGTH: 1266
```

```
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
```

```
; FEATURE:
```

```
; NAME/KEY: CDS
```

```
; LOCATION: (1)...(1266)
```

```
; US-09-993-059-11
```

Alignment Scores:

```
Prod. No.: 2,23e-249 Length: 1266
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
```

```
US-10-602-220-16 (1-417) x US-09-993-059-11 (1-1266)
```

```
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCCTCGGCC 60
QY 21 LeuValSerTyrAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGCTTTCTCTGGGACATCCCTGGGGCTAGAGCACTGGCAATGATTTGGCAAGGACGCT 120
QY 41 ThrMetGlyTyrLeuHisTyrGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTCAGAAAGCTCTTCATGGAGATGCGAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTyrMetAlaProGlnArg 100
Db 241 TGGAAAGGATCGAGTTATGAGTACCTCTGCATTTGATGACTGTTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyValArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAGGCGAGACTTCAGGCGAGCCCTCAGCGCTTTCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTTATGTTACACAGCAAGGACTGAAGCTAGGAGTATTATGCAGATGTTGGAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATACTACGACATTTGATGCCAGACCTTTGCT 480
QY 161 AspTyrGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAenLeu 180
Db 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTGTACTGTGACAGTTTGGAAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAenArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATTAAGCACATGCTCTGGCCCTGAATAGGACTGGCAGACATTTGTGTAC 600
QY 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
Db 661 CAGTACTGCAATCAGTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAAGTATAAAG 720
QY 241 SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGACCTGGACATCTTTTAACCCAGGAGAAATTTGATGTGTGACACCGGG 780
QY 261 GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGlnGln 280
```

```
Db      781  |||||GGTGGAGTACCCAGATATGTAGTATGGCAACTTTGGCTCAGCTGGATCAGCAA 840
Qy      281  ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db      841  GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTATTATCATGTCTAATGACCTC 900
Qy      301  ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db      901  CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGATAAGGACGTAATTCGCATCAAT 960
Qy      321  GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db      961  CAGGACCCCTTGGCAGCAAGGTTACCACTAGACAGGAGACAACTTTGAAGTGTG 1020
Qy      341  GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db      1021  GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATATAACCGCAGGAGATTGGT 1080
Qy      361  GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      1081  GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTCGGTAAGAGAGTGGCTGTAAATCCT 1140
Qy      381  AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db      1141  GCCTGCTTCATCACAGCTCTCTCTGTGAAGAGAGTAGGTTCTATGATGGACT 1200
Qy      401  SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db      1201  TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGTCTCAGCTA 1251
```

RESULT 4

```
US-10-103-327-11
; Sequence 11, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; PRIOR FILING DATE: 2002-03-20
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
```

LENGTH: 1266

TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1266)

US-10-103-327-11

Alignment Scores:

```
Pred. No.:      2,23e-249      Length:      1266
Score:          2293.00      Matches:      417
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     100.00%      Indels:      0
DB:              3              Gaps:      0
```

US-10-602-220-16 (1-417) x US-10-103-327-11 (1-1266)

```
Qy      1  MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db      1  ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTCCTCGCTTCCTCGGCC 60
Qy      21  LeuValSerTrpAspIleProGlyValaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
```

```
Db      61  CTCGTTTCTCTGGGACATCCCTCGGGCTAGAGCACTGGACAATGGATGGCAAGGACGCT 120
Qy      41  ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluPro 60
Db      121  ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCACACTTGACTGCCAGGAGAGCCA 180
Qy      61  AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db      181  GATTCTCTGCATCAGTGAGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAAGC 240
Qy      81  TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db      241  TGAAGAGATGCAAGTTATGAGTACCTTCGATTCATTCAGCTGTGGATGGCTCCCCAAGA 300
Qy      101  AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db      301  GATTTCAGAAGGAGAGCTTCAGGAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy      121  AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db      361  GCTAATTAATGTTACAGCAAGAGGACTGAAGCTAGGGAATTTATGCAGATGTGGAAATAA 420
Qy      141  ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db      421  ACCTGGCAGGCTTCCTCGGAGTTTTGGATATACGACATTCATGCCAGACCTTTGCT 480
Qy      161  AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db      481  GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGATTTGGAAATTTG 540
Qy      181  AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db      541  GCAGATGGTTATPAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGAGCATTTGTATC 600
Qy      201  SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db      601  TCCTGTGAGTGGCTCTTTATATATATGTGGCCCTTTCAAAAGCCCAATATATACAGAAATCCGA 660
Qy      221  GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db      661  CAGTACTGCAATCACCTGGCGAAATTTGCTGACATTCATGATGATTCCTGGGAAAGTATAG 720
Qy      241  SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db      721  AGTATCTTGGACTGGACATCTTTTAAACCAGAGAGAAATTTGTGATGTGCTGACCCAGGG 780
Qy      261  GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLysSerTrpAsnGlnGln 280
Db      781  GGTGGAAATGACCCAGATATGTTAGTATGGCACTTTTGGCCCTCAGCTGGAAATCAGCAA 840
Qy      281  ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db      841  GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGCTATGACCTC 900
Qy      301  ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db      901  CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGATAAGGACGCTAATTCGCATCAAT 960
Qy      321  GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db      961  CAGGACCCCTTGGGCAAGCAAGGTTACCACTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy      341  GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db      1021  GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATATAACCGCAGGAGATTGGT 1080
Qy      361  GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      1081  GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTCGGTAAGAGAGTGGCTGTAAATCCT 1140
Qy      381  AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db      1141  GCCTGCTTCATCACAGCTCTCTCTGTGAAGAGAGTAGGTTCTATGATGGACT 1200
```



```
; LOCATION: (1)...(1272)
US-10-103-327-17

Alignment Scores:
Pred. No.: 2,246-249 Length: 1272
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-16 (1-417) x US-10-103-327-17 (1-1272)

QY 1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTCTCTGGGC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTGGTTTCTTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGCTGGCTGCATCGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180

QY 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGlnLeuMetValSerGluGly 80
DB 181 GATTCTGCATCATGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240

QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGGAGGATGCAGGTTATGATGACCTCTGCAATGATGACTGTGGATGGCTCCCAAGA 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyLeuArgGlnLeu 120
DB 301 GATTTCAGAGGACGACTTCAGGCAGACCCCTCAGCGCTTCTCATGGATGCCAGCTA 360

QY 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyLeuTrpAlaAspValGlyAsnLys 140
DB 361 GCTAATATGTTCCACAGCAAGGACTGAAGCTAGGATTTATGCAGATGTGGAAATAA 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTyrAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATATCTACGATGATGCCAGACTTTGCT 480

QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGGAGTAGATCTGCTAAAAATTTGATGGTTGTACTGCACAGTTTGGAAAAATTG 540

QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 GCAGATGGTTATAAGCACATGTCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600

QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluLysArg 220
DB 601 TCTGTGAGTGGCTCTTTATATATGGCCCTTTCAAAGGCCAATATATACAGAAATCCGA 660

QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACTGGCGCAATTTTGTCTGACATGATGATCTTCTGGAAAGATATAAG 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTATGTGTGGACACAGGG 780

QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GGTGTGGANTGACCCAGATATGTTAGTATGTTGGCACTTTGGCCTCAGCTGGATCAGCAA 840

QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACATGAGTGGCCCTCTGGGCTCATATGGCTGCTCTTTATTCATGTCTAATGACCTC 900

QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACATCATCAGCCCTCAGCCAAAGCTCTCTTCCAGTAAGAGCTAATTTGCCATCAAT 960

QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTTGGGCAAGCAAGGTACCACTTACAGAGGAGAGCAACTTTGAAGTGTGG 1020

QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluLysGly 360
DB 1021 GAACGACCTCTCTCAGGCTTACCTTGGGCTGTAGCTATATATAACCGCAGGAGATTGGT 1080

QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1081 GGACCTGGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCCTGTAATCCT 1140

QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
DB 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAAGGAAGCTAGGGTTCTTATGAATGGACT 1200

QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
DB 1201 TCAAGGTTAAGAAGTCACATNAATCCACAGGACCTGTTTGTCTCAGCTA 1251

RESULT 7
US-09-993-059-7
; Sequence 7, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KIMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-09-993-059-7

Alignment Scores:
Pred. No.: 2,266-249 Length: 1278
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-16 (1-417) x US-09-993-059-7 (1-1278)

QY 1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTCTCTGGGC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTGGTTTCTTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGCTGGCTGCATCGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180

QY 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGlnLeuMetValSerGluGly 80
DB 181 GATTCTGCATCATGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
```

QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGGATGAGGTTATGAGTACCTCTGCAATTGATGACTGTGTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGAGAGCTTCAGGAGAGACCTTCAGGCGCTTCTCTATGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTTCACAGCAAGGAGCTGAAGCTAGGAGTTATGCAGATGTGGGAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGGCTTCCTGGGAGTTTGGATACCTAGCAGATTCATGCCAGACCTTTGCT 480
QY 161 AspTyrGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGACAGTTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGCACATGCTCTTGGCCCTGAATAGACTGGCAGAGCATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCAATTGATGATTCCTGGAAAGTATAAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnAlaGlyIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGACATGGACATCTTTTACCAGAGAGAAATTTGTGATGTGTGACCCAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTGGTGGCACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCATGCCCTCAGCCAAAGCTCTCTTCAGGATAGGACGTAATTCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGAGCAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCGCTGGCTGTAGCTATGATAAACCGCGAGGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTTCGCTCTTATACCATCGAGTTGCTTCTCCCTGGTAAAGAGTGGCCTGTAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTGCTTCAGCTA 1251

RESULT 8

US-10-103-327-7

; Sequence 7, Application US/10103327

; Patent No. 6890748

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-10-103-327-7

Alignment Scores:

Score: 2,266-249 Length: 1278
2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-16 (1-417) x US-10-103-327-7 (1-1278)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTCAGCAAGCAACCCAGCACTACATCTGGGCTGGCGCTTGGCTTCTCGCTCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCGGACATCCCTGGGGCTAGAGCAGCTGGACAAATGGATTTGGCAAGCGCCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACTTGCCTGACCCAGGAAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGTCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGGATGAGGTTATGAGTACCTCTGCAATTGATGACTGTGTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGAGAGCTTCAGGAGACCTTCAGCGCTTCTCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTTCACAGCAAGGAGCTGAAGCTAGGAGTTATGCAGATGTGGGAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGGCTTCCTGGGAGTTTGGATACCTAGCAGATTCATGCCAGACCTTTGCT 480
QY 161 AspTyrGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGACAGTTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGCACATGCTCTTGGCCCTGAATAGACTGGCAGAGCATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220

Db 1021 GAACGACCTCTCTCAGGCTTAGGCTGGCTGTAGCTATGATAAACCGGAGGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTATACCATCGAGTGTCTTCCCTGGGTAAAGAGTGGCTGTATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGACTCCTCCCTGTGAAAGGAAGCTAGGGTTCATGATGAGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTACATAAATCCACAGGCACTGTTTTGCTTCAGCTA 1251

RESULT 10
US-10-103-327-13
; Sequence 13, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1284)
US-10-103-327-13

Alignment Scores:
Pred. No.: 2,27e-249 Length: 1284
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-16 (1-417) x US-10-103-327-13 (1-1284)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGACCCAGACTACATCTGGGCTGGCGCTGGCGCTTCGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGCAATGATTGGCAAGGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGTGGCAGAGCTCATGTCTCAGAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGAAGGATGCAGTTATCAGTACCTCTGCAATTGATGACTGTGTGGATGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120

Db 301 GATTCAAGAGGAGCAGACTTCAGGAGAGCCCTCAGCGCTTCTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTTACACAGCAAGGAGCTGAAGCTAGGATTTATGCAGATGTTGGAAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTTGGATATCTACGACATTTGATGCCAGACTTTGCT 480
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTTACTGTGACACTTTGGAAATTTG 540
QY 181 AlaAspGlyTyrIleHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCACTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLleArg 220
Db 601 TCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAGGCCCAATTTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCATACACTGGCGAAATTTGCTGACATTTGATGATTCCTGGAAAGATATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAACACAGGAGAGAAATTTGTTGATGCTTGTGACCCAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGATGACCCAGATATGTTAGTATGGCACTTTGGCCCTCAGCTGGATACGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACCACTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAACCGGAGGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGAGTGTCTTCCCTGGGTAAAGAGTGGCTGTATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCATGATGAGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTACATAAATCCACAGGCACTGTTTTGCTTCAGCTA 1251

RESULT 11
US-09-993-059-3
; Sequence 3, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06


```

; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1290)
US-09-993-059-3

Alignment Scores:
Pred. No.: 2,29e-249 Length: 1290
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-16 (1-417) x US-09-993-059-3 (1-1290)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCGCTTCCTCGCC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTTCCTGGGACATCCCTGGGCTAGACACTGGACATGGATGGCAGGAGCGCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGCTGGCTGCTGCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCA 180

QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCCTGCATCAGTCAGGAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240

QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGAAGGATGCAGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTCAGAGGCAGACTTCAGGCGAGCCCTCAGCGCTTTCCTCATGGATTGCCAGCTA 360

QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 361 GCTAATATGTTACACAGCAAGGACTGAAGCTAGGATTTATGATGATGATGATGATGAT 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCGAGCTTCCTCGGAGTTTGGATATCTACGACATTTGATGCCAGAGCTTTGCT 480

QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGGAGTAGATCTGCTAATAATTTGATGTTTACTGTGACAGTTTGGAAAAATTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 GCAGATGGTTATAGCACATGCTCCTGGCCCTGAAATAGGACTGGCAGAGCATTTGTGTAC 600

QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
DB 601 TCTGTGAGTGGCCCTCTTTATATATGTCCTTTCCTTCAAAAGGCCAATTATACAGAAATCCGA 660

QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCATCATCTGCGGAAATTTTGTGCTGACATTGATGATTCCTGGAAAAAGTATAAG 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCGCTTCCTCGCC 60

721 AGTATCTGGACTGGACATCTTTTAAACCCAGCAGAGAAATGTTGATGTTTGGACCGAGG 780
261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
781 GGTGGATGACCCAGATATGTAGTAGTATGGCACTTTGGCCTCAGCTGGATCAGCA 840
281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTATTCATGCTTAATGACCTC 900
301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGAGCTAATTCCTCATCAAT 960
321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
961 CAGGACCCCTTTGGGCAAGGATGACGCTTAGACAGGGAGACAACCTTTGAAAGTGG 1020
341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGAGCTATGATATAAACCAGCAGGAGATTGCT 1080
361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAATCCT 1140
381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAGGAGCTAGGCTTCTATGATGGACT 1200
401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
1201 TCAAGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCCTTCAGCTA 1251

RESULT 12
US-10-103-327-3
; Sequence 3, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1290)
US-10-103-327-3

Alignment Scores:
Pred. No.: 2,29e-249 Length: 1290
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-16 (1-417) x US-10-103-327-3 (1-1290)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCGCTTCCTCGCC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTTCCTGGGACATCCCTGGGCTAGACACTGGACATGGATGGCAGGAGCGCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGCTGGCTGCTGCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCA 180

QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCCTGCATCAGTCAGGAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240

QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGAAGGATGCAGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTCAGAGGCAGACTTCAGGCGAGCCCTCAGCGCTTTCCTCATGGATTGCCAGCTA 360

QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 361 GCTAATATGTTACACAGCAAGGACTGAAGCTAGGATTTATGATGATGATGATGATGAT 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCGAGCTTCCTCGGAGTTTGGATATCTACGACATTTGATGCCAGAGCTTTGCT 480

QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGGAGTAGATCTGCTAATAATTTGATGTTTACTGTGACAGTTTGGAAAAATTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 GCAGATGGTTATAGCACATGCTCCTGGCCCTGAAATAGGACTGGCAGAGCATTTGTGTAC 600

QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
DB 601 TCTGTGAGTGGCCCTCTTTATATATGTCCTTTCCTTCAAAAGGCCAATTATACAGAAATCCGA 660

QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCATCATCTGCGGAAATTTTGTGCTGACATTGATGATTCCTGGAAAAAGTATAAG 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCGCTTCCTCGCC 60
```


Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCCTGGGACATCCCTGGGGCTAGAGCACTGGACAATGATGGCAAGGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTTCCTGCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGATGACGGTTATGATGACTCTCTGCAATGATGACTGTTGGATGGCTCCCAAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGCAGACTTCAGGCAAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
Db 361 GCTAATATTGTTACAGCAAGGACTCAAGCTAGGAGTTTATGCAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGCTTCCTGGGAGTTTGGATACCTACGACATGATGCCAGCCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGACTGCTAAATTTGATGGTGTGTACTGTGACAGTTTGGAAATTTG 540
Qy 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
Db 541 GCAGATGTTTATAGCACATGCTCTTGGCCCTGANTAGACTGGCAGAGCATTTGTATC 600
Qy 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Qy 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCAATTGATGATTCCTGGAAAGATATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAACCAAGAGAGAAATTTGTGATGTTGCTGACCAAGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTGTGATGGCACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATAGCCCTCAGGCAAGCTCTCTTCAGGATAAGCAGCTAATGTCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGCAAGCAAGGTTACCGCTTAGACAGGAGACACAACCTTGAAGTGTG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCGCTGGCTGTAGCTATGATAAACCGGAGGAGATTGCT 1080
Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCTCCCTGGTAAAGAGTGGCCTGTAATCT 1140

Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
Db 1141 GCCTGTCTTCATCACAGACTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGATGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTACATAAAATCCACAGGCACCTGTTTTGCTTCAGCTA 1251
RESULT 13
US-09-993-059-9
; Sequence 9, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1296
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1296)
US-09-993-059-9
Alignment Scores:
Pred. No.: 2,31e-249 Length: 1296
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
US-10-602-220-16 (1-417) x US-09-993-059-9 (1-1296)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTCAGGAACCCAGAACCTACATCTGGGCTGGCGCTTGGCGCTTGGCTTCTCGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCCTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATGGCAAGGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTTCCTGCATCAGTGAAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGATGACGGTTATGATGACTCTTCATGATGATGATGATGATGATGATGATGATGAT 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGCAGACTTCAGGCAAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
Db 361 GCTAATATTGTTACAGCAAGGACTCAAGCTAGGAGTTTATGCAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGCTTCCTGGGAGTTTGGATACCTACGACATGATGCCAGCCTTTGCT 480

		; ORGANISM: Homo sapiens	
		; FEATURE:	
		; NAME/KEY: CDS	
		; LOCATION: (1)...(1296)	
		US-10-103-327-9	
		Alignment Scores:	
		Pred. No.: 2,31e-249	Length: 1296
		Score: 2293.00	Matches: 417
		Percent Similarity: 100.00%	Conservative: 0
		Best Local Similarity: 100.00%	Mismatches: 0
		Query Match: 100.00%	Indels: 0
		DB: 3	Gaps: 0
		US-10-602-220-16 (1-417) x US-10-103-327-9 (1-1296)	
QY	1	MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20
DB	1	ATGCAGCTGAGGAAACCAGAACTACATCTGGGCTGGCGCTTGGCGCTTCCTTCCCTGGCC	60
QY	21	LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro	40
DB	61	CTCGTTTCTCTGGGACATCCCTGGGGCTAGAGCACCTGGCAATGGATTGGCAAGAGCGCT	120
QY	41	ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro	60
DB	121	ACCATGGGCTGGCTGCATGGGAGGGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA	180
QY	61	AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly	80
DB	181	GATTCTGCATCAGTGAGAAAGCTTTCATGGAGATGGCAGAGCTCATGTCTCTCAGAAGGC	240
QY	81	TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg	100
DB	241	TGGAAGGATGCAAGTTATGAGTACCTTCGATTCGATGACTGTGTGGATGGCTCCCCAAGA	300
QY	101	AspSerGluValArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu	120
DB	301	GATTTCGAAGGACAGACTTCAGGCAGACCTTCAGGGCTTTCTCTCATGGATTCGCCAGCTA	360
QY	121	AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys	140
DB	361	GCTAATATATGTTTACAGCAAGAGACTTGAAGCTAGGAGTTTATGCAGATGTGGAAATAAA	420
QY	141	ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla	160
DB	421	ACCTGGCGAGGCTTCCTCTGGGAGTTTGGATATCTACGACATTCATGCCAGACCTTTGCT	480
QY	161	AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu	180
DB	481	GACTGGGAGTAGATCTGCTTAAATTTGATGGTTGTTACTGTGCAGTTTGGAAAAATTG	540
QY	181	AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr	200
DB	541	GCAGATGGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAACATTGTGTAC	600
QY	201	SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg	220
DB	601	TCCTGTGAGTGGCTCTCTTTATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA	660
QY	221	GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys	240
DB	661	CAGTACTGCAATCAGTGGCGAATTTTGTCTGACATTCATGATTCCTCGGAAAAATATAAG	720
QY	241	SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly	260
DB	721	AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTGTGTATGTGTGTGGACCAAGG	780
QY	261	GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln	280
DB	781	GGTTGGAAATGACCCAGATATGTAGTGTATGGCACTTTGGCCCTCAGCTGGAATCAGCAA	840
QY	281	ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu	300
DB	841	GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTATGTCCTAATGACCTC	900
QY	301	ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn	320
DB	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGGCCATCAAT	960
QY	321	GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp	340
DB	961	CAGGACCCCTTGGGCAAGCAGGGTACCGACTTAGACAGGAGACAACTTTGAAGTGTGG	1020
QY	341	GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly	360
DB	1021	GAACGACCTCTCTCAGGCTTAGCGCTGGCTGTAGCTATGATAAACCGCGCAGGAGATTGGT	1080
QY	361	GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro	380
DB	1081	GGACCTCGCTCTTATACCAATCGCAGTTCCTTCCCTGGGTAAAGGAGTGGCTGTAACTCT	1140
QY	381	AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr	400
DB	1141	GCCTGCTTTCATCACACAGCTCTCTCCCTGTGAAAAGGAGCTAGGGTTCTATGAATGGACT	1200
QY	401	SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu	417
DB	1201	TCAAGGTTAAGAGTCACATAAATCCACAGGCAGCTGTTTGTCTTCAGCTA	1251
RESULT 14			
US-10-103-327-9			
; Sequence 9, Application US/10103327			
; Patent No. 6890748			
; GENERAL INFORMATION:			
; APPLICANT: GARGER, Stephen A.			
; APPLICANT: TURPEN, Thomas H.			
; APPLICANT: KUMAGAI, Monto H.			
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN			
; PRIOR INVENTION: PLANTS BY TRANSIENT EXPRESSION			
; FILE REFERENCE: 008010087CPUS06			
; CURRENT APPLICATION NUMBER: US/10/103,327			
; CURRENT FILING DATE: 2002-03-20			
; PRIOR APPLICATION NUMBER: US/09/993,059			
; PRIOR FILING DATE: 2001-11-13			
; NUMBER OF SEQ ID NOS: 37			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 9			
; LENGTH: 1296			
; TYPE: DNA			

Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCATGACCTCAAGCCAAAGCTCTCTTCAGATTAAGACGTAATGTCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTAACAGCTAGACAGGAGACAACCTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGGCTGGCTGTAGCTATGATAAACCCGCGAGGAGATGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTATACCATCGCAGTGTCTTCCCTGGGTAAAGAGATGGCTGTATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrrpThr 400
Db 1141 GCCTGTCTCATCACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTCTTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTCAATAAATCCACAGGCACTGTTTGTCTCAGCTA 1251

RESULT 15

US-09-993-059-5
; Sequence 5, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KIMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1308
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1308)
US-09-993-059-5

Alignment Scores:
Pred. No.: 2,34e-249 Length: 1308
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-16 (1-417) x US-09-993-059-5 (1-1308)

Qy 1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTGCGCTTCTCGCTTCTCGGCC 60
Qy 21 LeuValSerTrrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTrrpLeuHisTrrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGGGAGCGCTTCATGTGCACCTTGTACTGCCAGGAAGAGCCA 180

Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTTCCTGCATCAGTCAGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTrrpLeuCysIleAspAspCysTrrpMetAlaProGlnArg 100
Db 241 TGAAGGATGCGAGTTATGAGTACCTCTGCATTGATGACTGTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGAGCAGACTTCAGGAGACCTTCAGCGCTTCTCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTrrpValHisSerLysGlyLeuLysLeuGlyIleTrrpAlaAspValGlyAsnLys 140
Db 361 GCTAATATATGTTTACACAGCAAGGACTGAAGCTAGGAGTTATGACAGATGTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrrpTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTAGCAGATTGATGCCAGACTTGTGCT 480
Qy 161 AspTrrpGlyValAspLeuLysPheAspGlyCysTrrpCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGACTGCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTG 540
Qy 181 AlaAspGlyTrrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrr 200
Db 541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGACATGTGTAC 600
Qy 201 SerCysGluTrrpProLeuTrrpMetTrrpProPheGlnLysProAsnTrrpThrGluIleArg 220
Db 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATATACAGAAAATCCGA 660
Qy 221 GlnTrrpCysAsnHisTrrpArgAsnPheAlaAspIleAspAspSerTrrpLysSerIleLys 240
Db 661 CAGTACTGCAATCAGCTGGCGAATTTTGTGTCATGTGATGATTCCTGGAAAAGTATAAG 720
Qy 241 SerIleLeuAspTrrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTAAACACAGGAGAGAAATTTGTGATGTCTGCGACAGGG 780
Qy 261 GlyTrrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrrpAsnGln 280
Db 781 GGTGGATGACCCAGATATGTAGTGTGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTATTCATGTCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCATGACCTCAAGCCAAAGCTCTCTTCAGGATAAGACGTAATTCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTAACAGCTAGACAGGAGACAACCTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATAAACCCGCGAGGAGATGGT 1080
Qy 361 GlyProArgSerTrrpIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGGAGTGGCTGTATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrrpThr 400
Db 1141 GCCTGTCTCATCACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTCTTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTCAATAAATCCACAGGCACTGTTTGTCTCAGCTA 1251

Tue Jan 3 11:36:07 2006

Search completed: January 1, 2006, 05:53:05
Job time : 367.352 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2006, 00:27:17 ; Search time 7564.48 Seconds
(without alignments)
2579.189 Million cell updates/sec

Title: US-10-602-220-16

Perfect score: 2293
Sequence: 1 MQLRNPELHGCALALRFLA.....EWTSLRSHINPTGTLLQL 417

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO_spool/US1062220/runat_30122005_140558_5433/app_query.fasta_1.782
-DB=EST -QFMT=fastcap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US1062220@cgn 1.1 6731 @runat_30122005_140558_5433 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

```
EST:*
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2293	100.0	1266	4 CR607242	CR607242 full-length
2	2293	100.0	1277	4 CR605654	CR605654 full-length
3	2289	99.8	1253	4 CR617861	CR617861 full-length
4	2083	90.8	1290	10 AY408540	AY408540 Homo sapi
5	2034	88.7	1290	10 AY408541	AY408541 Pan trogl
6	1806	78.8	1376	4 AK054547	AK054547 Mus muscu
7	1806	78.8	2962	4 AK040081	AK040081 Mus muscu

8	1777	77.5	1086	1	AL554978	AL554978
9	1699	74.1	1014	5	EX354096	EX354096
10	1691.5	73.8	999	1	AL552630	AL552630
11	1667.5	72.7	1067	3	BM564282	BM564282
12	1659.5	72.4	1005	3	BQ062192	AGENCOURT
13	1649	71.9	922	7	CO645623	ILLUMIGEN
14	1648.5	71.9	1026	1	AL577581	AL577581
15	1627	71.0	1296	10	AY408542	AY408542 Mus muscu
16	1620.5	70.7	1133	7	CO645672	ILLUMIGEN
17	1589	69.3	920	7	CO645464	ILLUMIGEN
18	1578	68.8	923	7	CT005156	CT005156
19	1574	68.5	984	7	CO646251	ILLUMIGEN
20	1556.5	67.9	976	5	BQ956043	AGENCOURT
21	1527	66.6	898	5	BU191867	AGENCOURT
22	1519	66.2	922	5	BU540848	AGENCOURT
23	1515.5	66.1	937	5	BQ934640	AGENCOURT
24	1492.5	65.1	852	6	CA454083	AGENCOURT
25	1465	63.9	888	5	BU154569	AGENCOURT
26	1458	63.6	958	6	CA487415	AGENCOURT
27	1452	63.3	871	6	CA454143	AGENCOURT
28	1443	62.9	801	5	BU596617	AGENCOURT
29	1429	62.3	1025	3	BM450649	AGENCOURT
30	1425.5	62.2	849	6	CA487531	AGENCOURT
31	1422.5	62.0	879	2	EG824387	AGENCOURT
32	1412.5	61.6	927	2	BE622583	AGENCOURT
33	1408.5	61.4	871	3	BQ225444	AGENCOURT
34	1406	61.3	908	2	BE379425	AGENCOURT
35	1404	61.2	796	8	DR156332	HESC2_64
36	1402.5	61.2	921	6	CA488935	AGENCOURT
37	1397	60.9	845	7	CR983967	CR983967
38	1395.5	60.9	1071	1	AL575861	AL575861
39	1384	60.4	813	5	EX344841	EX344841
40	1382	60.3	782	2	B1224248	602940538
41	1371.5	59.8	1066	2	BG824323	602727346
42	1365	59.5	749	5	EX344842	EX344842
43	1360	59.3	908	6	CD385342	AGENCOURT
44	1326	57.8	886	3	BP435628	BP435628
45	1303	56.8	767	3	BI753664	BI753664

ALIGNMENTS

RESULT 1

CR607242

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1266 bp mRNA linear HTC 21-JUL-2004
full-length cDNA clone CSODI067VJ01 of Placentia Cot 25-normalized
of Homo sapiens (human).

CR607242 GI:50488049

HTC; CNSLT_CDNA.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 1266)

Li, W.B., Gruber, C., Jesse, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished

Contact: Feng Liang Email: fliang@lifetech.com URL:

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue

2 (bases 1 to 1266)

Genoscope.

Direct Submission

Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage:

BP 191 91006 EVRY cedex - FRANCE (E-mail: seque@genoscope.cns.fr

- Web: www.genoscope.cns.fr)

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen.

FEATURES		Location/Qualifiers	
source		1..1266	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clones="CS0D1067YJ01"	
		/tissue_type="Placenta Cot 25-normalized"	
		/plasmid="pCMVSPORT_6"	
ORIGIN			
Alignment Scores:			
Pred. No.:		1.21e-240	
Score:		2293.00	
Percent Similarity:		100.00%	
Best Local Similarity:		100.00%	
Query Match:		100.00%	
DB:		4	
US-10-602-220-16 (1-417) x CR607242 (1-1266)			
QY	1	MetGlnLeuArgAnpProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20
DB	7	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCGCTTCCTGGCC	66
QY	21	LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAnGlyLeuAlaArgThrPro	40
DB	67	CTCGTTTCTGGGACATCCCTGGGGCTTAGACACTGGACATGGATTGGCAAGGACGCT	126
QY	41	ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAnLeuAspCysGlnGluPro	60
DB	127	ACCATGGGCTGGCTGCATCGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA	186
QY	61	AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly	80
DB	187	GATTCTCTGCATCAGTGAGAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAGGC	246
QY	81	TrpLysAspAlaGlyTrpGluTrpLeuLysLeuGlyCysIleAspAspCysTrpMetAlaProGlnArg	100
DB	247	TGGAGGATGACGCTTATGATGACTCTGCAATTGATGATGATGGTGGCTCCCAAGA	306
QY	101	AspSerGluArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu	120
DB	307	GATTCAAGAGCAGACTTCAGGCAGACCCCTCAGCGCTTCTCATGGGATTGCCAGCTA	366
QY	121	AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys	140
DB	367	GCTAATTATGTTTCACCAAGAGGACTGAAGTAGGATTTATGAGATGTTGGAAATAA	426
QY	141	ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla	160
DB	427	ACCTGGCAGGCTTCCCTGGGAGTTTGGATCTACGACATTGATGCCAGACCTTTGCT	486
QY	161	AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu	180
DB	487	GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG	546
QY	181	AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp	200
DB	547	GCAGATGGTTATAAGCAGCATGCTCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC	606
QY	201	SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg	220
DB	607	TCTGTGAGTGGCTCTTTATATGTGGCCCTTCCAAAGGCCAATTATACGAATTCGA	666
QY	221	GlnTrpCysAnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys	240
DB	667	CAGTACTGCAATCACTGGCGAAATTTTGTGCTGACATTGATGATTCCTGCAAAAGTATAAG	726
QY	241	SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly	260
DB	727	AGTATCTTGACCTGGAATCTTTTAAACAGAGAGAAATTGTGTGCTGGACACAGGG	786
QY	261	GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln	280
Alignment Scores:			
Pred. No.:		1.21e-240	
Score:		2293.00	
Percent Similarity:		100.00%	
Best Local Similarity:		100.00%	
Query Match:		100.00%	
DB:		4	
US-10-602-220-16 (1-417) x CR607242 (1-1266)			
QY	1	MetGlnLeuArgAnpProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20
DB	7	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCGCTTCCTGGCC	66
QY	21	LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAnGlyLeuAlaArgThrPro	40
DB	67	CTCGTTTCTGGGACATCCCTGGGGCTTAGACACTGGACATGGATTGGCAAGGACGCT	126
QY	41	ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAnLeuAspCysGlnGluPro	60
DB	127	ACCATGGGCTGGCTGCATCGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA	186
QY	61	AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly	80
DB	187	GATTCTCTGCATCAGTGAGAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAGGC	246
QY	81	TrpLysAspAlaGlyTrpGluTrpLeuLysLeuGlyCysIleAspAspCysTrpMetAlaProGlnArg	100
DB	247	TGGAGGATGACGCTTATGATGACTCTGCAATTGATGATGATGGTGGCTCCCAAGA	306
QY	101	AspSerGluArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu	120
DB	307	GATTCAAGAGCAGACTTCAGGCAGACCCCTCAGCGCTTCTCATGGGATTGCCAGCTA	366
QY	121	AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys	140
DB	367	GCTAATTATGTTTCACCAAGAGGACTGAAGTAGGATTTATGAGATGTTGGAAATAA	426
QY	141	ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla	160
DB	427	ACCTGGCAGGCTTCCCTGGGAGTTTGGATCTACGACATTGATGCCAGACCTTTGCT	486
QY	161	AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu	180
DB	487	GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG	546
QY	181	AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp	200
DB	547	GCAGATGGTTATAAGCAGCATGCTCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC	606
QY	201	SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg	220
DB	607	TCTGTGAGTGGCTCTTTATATGTGGCCCTTCCAAAGGCCAATTATACGAATTCGA	666
QY	221	GlnTrpCysAnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys	240
DB	667	CAGTACTGCAATCACTGGCGAAATTTTGTGCTGACATTGATGATTCCTGCAAAAGTATAAG	726
QY	241	SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly	260
DB	727	AGTATCTTGACCTGGAATCTTTTAAACAGAGAGAAATTGTGTGCTGGACACAGGG	786
QY	261	GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln	280

```
Pred. No.: 1.22e-240 Length: 1277
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-602-220-16 (1-417) x CR605654 (1-1277)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 10 ATGCAGCTGAGGACCCAGAACTACATCTGGCTGGCGCTTGGCTTCTTCTGGCC 69
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgTrpPro 40
Db 70 CTGCTTCTGGGACATCCCTGGGCTAGACACTGGACAAATGGATTGGCAAGACGCT 129
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 130 ACCATGGGCTGGCTGACACTGGGAGCGCTTTCATGTGCAACCTTGACTGCCAGAAAGGCCA 189
Qy 61 AspSerCysIleSerGluLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 190 GATTCCTGCATCAGTGAAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 249
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 250 TGGGAAGATGAGGTTATGAGTACCTCTGCATTCATGACTCTGGATGGCTCCCAAGA 309
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 310 GATTCAGAAGCGACACTTCAGGAGAGCCCTCAGGCGCTTTCATGAGGATTCGCCAGCTA 369
Qy 121 AlaSerTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
Db 370 GCTAATATGTTACAGCAAGGACTGAGTAGGATTTATGAGATTTGGAAATATA 429
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnTrpPheAla 160
Db 430 ACCTGGCAGGCTTCCCTGGGAGTTTGGTACTACGACATTTGATGCCAGACCTTTGCT 489
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db 490 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACGATTTGGAAATTTG 549
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
Db 550 GCAGATGTTATAGCACATCTCTTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC 609
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTrpThrGluLeuArg 220
Db 610 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 669
Qy 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 670 CAGTACTGCAATCACTGGCGAAATTTTGTGCTACATTTGATGATTCCTGAAAGATATAAG 729
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 730 AGTATCTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTGATTTGTGTCGACAGGG 789
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 790 GGTGGGAATGACCAGATATGTTAGTATGTTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 849
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 850 GTAACTCAGATGGGCTCTGGGCTATCATGCTCTCTCTTTATTCATGCTCTTAATGACCTC 909
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 910 CGACACATCAGCCCTCAGCCCAAGCTCTCTTCAGGATAGGACGTAATTTGCCATCAAT 969
```

```
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 970 CAGGACCCCTTGGCAAGCAAGGTTACCAAGCTTAGACAGGAGAGCAACTTTTGAAGTGTGG 1029
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1030 GACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCCGCGAGGATTTGT 1089
Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1090 GGACCTCGCTTATACCATCGCAGTGTCTCCCTGGTAAAGGAGTGGCTGTATATCT 1149
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1150 GCCTGTCTTCATCACACAGCTCTCCCTGTGAAAAGGAAGCTAGGTTCTATGAATGGACT 1209
Qy 401 SerArgLeuArgSerHisLeuAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1210 TCAAGTTAGAGTACATATAATCCACAGGCACTGTTTGTCTTCAGCTA 1260

RESULT 3
CR617861 1253 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CSODK009YE12 of HeLa cells Cot 25-normalized
of Homo sapiens (human).
ACCESSION CR617861
VERSION CR617861.1 GI:50498668
KEYWORDS HTC; CNSLT_cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 1253)
AUTHORS Li, W.B., Gruber, C., Jesses, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue
REFERENCE 2 (bases 1 to 1253)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life technologies, a
division of Invitrogen.
FEATURES
source
1..1253
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODK009YE12"
/tissue types="HeLa cells Cot 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN
Alignment Scores: 3.27e-240 Length: 1253
Pred. No.: 2289.00 Matches: 416
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.83% Indels: 0
DB: 4 Gaps: 0

US-10-602-220-16 (1-417) x CR617861 (1-1253)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
```

```
Db      6  ATGCAGCTGAGGAACCCAGAACTACATCTGGCGTGGCGCTTGGCTTGGCTTCTCGGCC 65
Qy      21  LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db      66  CTCGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGACGCCT 125
Qy      41  ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db      126  ACCATGGGCTGGCTGGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGACCA 185
Qy      61  AspSerCysIleSerGluLeuLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db      186  GATTCTCGCATCACTGAGAGCTCTTCATGAGATGCGAGAGCTCATGCTCAGAAAGGC 245
Qy      81  TrpIysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db      246  TGGAGGATGCAGGTTATGACTGCTCTGCACTTGCATGATGACTGTGGATGGCTCCCAAGA 305
Qy      101  AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db      306  GATTTCAGAAGCGACATTCAGCGAGACCTCAGCGCTTTCCTCATGGGATTCGCAGCTA 365
Qy      121  AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
Db      366  GCTAATTATGTTACAGCAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAAATAAA 425
Qy      141  ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
Db      426  ACCTGCGCAGGCTTCCCTGGGAGTTTGGTACTACGACATTCATGATGCGCAGACCTTGGCT 485
Qy      161  AspTrpGlyValAspLeuLeuPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db      486  GACTGGGAGTAGATCTGCTAATAATTGATGGTTTACTCTGACAGTTTGGAAAAATTG 545
Qy      181  AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
Db      546  GCAGATGTTTATAAGCACATGCTCTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTATC 605
Qy      201  SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
Db      606  TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGGCCAATTATACAGAAATCCGA 665
Qy      221  GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db      666  CAGTACTGCATCACTGGCGAAATTTTCTGACATTTGATGATCTCTGGAAAGATATAAG 725
Qy      241  SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db      726  AGTATCTTGGACTGGACATCTTTTAAACCAGGAGAGAAATTGTTGATGTTGCTGGACAGG 785
Qy      261  GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db      786  GGTTTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTTGGCCCTCAGCTGGAAATCAGCAA 845
Qy      281  ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
Db      846  GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTTTTATTCATGCTTAATGACCTC 905
Qy      301  ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db      906  CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAAGGACGTAATTTGCCATCAAT 965
Qy      321  GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db      966  CAGGACCCCTTTGGCAAGCAAGGGTACCAGCTTATGACAGGGAGACAATTTTGAAGTGTGG 1025
Qy      341  GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db      1026  GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATTAACCCGCGCAGAGATTGGT 1085
Qy      361  GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      1086  GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1145
```

```
Qy      381  AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
Db      1146  GCTGCTGTTCAATCACACAGCTCCCTCCCTGTGAAAAGGAGTAGGTTCATGATGAGCT 1205
Qy      401  SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGln 416
Db      1206  TCAAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTGTCTTCTCAG 1253

RESULT 4
AY408540      1290 bp      DNA      linear      GSS 15-DEC-2003
LOCUS      Homo sapiens GLA gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION      AY408540
VERSION      AY408540.1 GI:39764511
SOURCE      GSS.
ORGANISM      Homo sapiens (human)
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE      1 (bases 1 to 1290)
AUTHORS      Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Perriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
TITLE      Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
JOURNAL      Science 302 (5652), 1960-1963 (2003)
PUBMED      14671302
REFERENCE      2 (bases 1 to 1290)
AUTHORS      Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Perriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
TITLE      Direct Submission
JOURNAL      Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT      This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES             Location/Qualifiers
            source             1..1290
                                /organism="Homo sapiens"
                                /mol_type="genomic DNA"
                                /db_xref="taxon:9606"
            gene              <1..>1290
                                /gene="GLA"
                                /locus_tag="HCM3258"
            ORIGIN
Alignment Scores:      1.51e-217      Length:      1290
Pred. No.:            2083.00      Matches:      386
Score:
Percent Similarity:    92.57%      Conservative: 0
Best Local Similarity: 92.57%      Mismatches:   31
Query Match:          90.84%      Indels:       0
DB:                    10      Gaps:         0

US-10-602-220-16 (1-417) x AY408540 (1-1290)
Qy      1  MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db      1  ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTCGGCC 60
Qy      21  LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db      61  CTCGTTTCTGGGACATCCCTGGGCTTAGACACTGGGCAATGGATTGGCAAGACGCCT 120
Qy      41  ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db      121  ACCATGGGCTGGCTGGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGACCA 180
```



```
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAGGATGAGGTTATGAGTACTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTGAGAGGACACTTCAGGCAGACCTTCAGCGCTTCTCTCATGGATTGGCCAGCTA 360
Qy 121 AlaSerTyrValHisSerLysGlyLeuLysGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTTATGTTCAAGCAAGAGTCAAGCTAGGATGATGATGAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATGACTTACGACATTTGATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAATTTTGGTACATTTGATGATGATGATGATGATGAT 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAATTTGATGTTGCTGGACACGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db 781 GGTGGAAATGACCAGATATGTTAGTATGATGGCAACTTTGGCCCTCAGCTGAATCAGCA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTATTCATGCTCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGCAAGCAAGGTTACAGCTTAGACAGGAGAGACAACTTTGAAGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCGGAGGATTTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaLysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTGTAAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAGGAGCTAGGGTTCTATGAATGAGCT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAGTCACTAAATCCACAGGCACTGTTTGTCTTCAAGCTA 1251
```

RESULT 5

```
AY408541 1290 bp DNA linear GSS 15-DEC-2003
LOCUS Pan troglodytes GLA gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY408541
VERSION AY408541.1 GI:39764512
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Pan.
REFERENCE 1 (bases 1 to 1290)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1290)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source Location/Qualifiers
1..1290
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1290
/gene="GLA"
/locus_tag="HCM3258"
gene
ORIGIN
Alignment Scores: 3.66e-212 Length: 1290
Pred. No.: 2034.00 Matches: 377
Score: 2034.00
Percent Similarity: 90.89% Conservative: 2
Best Local Similarity: 90.41% Mismatches: 38
Query Match: 88.70% Indels: 0
DB: 10 Gaps: 0
US-10-602-220-16 (1-417) x AY408541 (1-1290)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGCTGCGCTTGGCTTCGCTTCCTGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCTTTTCTGGGACATCCCTTGGGCTTAGACACTGAGCAATGGATTGGCAAGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCTATGTGCAACCTTGACTGCTCCAGGAAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAGGATGAGGTTATGAGTACTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
```

```
Db      301  GATTTCAGAGCGAGACTTCAGGCAGACCCCTCAGCGCTTTCTCTCATGGGATTCGNNAGCTA 360
Qy      121  AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db      361  NNNNNNNNGTTTCACAGCAAGGACTGAAGTAGGGATTTATGCAGATGTTGGAAATAAA 420
Qy      141  ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db      421  ACCTGCCGAGCTTCCCTCGGAGTTTTGGTACTACGACATTTGATGCCAGACCTTTGCT 480
Qy      161  AspTyrGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db      481  GACTGGGAGTAGATCTGCTAAATTTTGATGTTGTTACTGTGCACAGCTTTGGAAAATTTG 540
Qy      181  AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db      541  GCAGATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 600
Qy      201  SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
Db      601  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 660
Qy      221  GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
Db      661  CAGTACTGCAATCACTCGCGAAATTTTGTGCTGACCTTGATGATTCCTCGAAAAGTATAAG 720
Qy      241  SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db      721  AGTATCTTGGACTGGACATCTTTTAAACGAGAGAGAAATTTGTGATGTTGCTGGACCAAGG 780
Qy      261  GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGlnGln 280
Db      781  GGTTGGAAATGACCCAGATATGTTAGTATGGCAACTTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy      281  ValThrGlnMetAlaLeuTyrAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db      841  GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTCTCTTTTATTCATGCTTAATGACCTC 900
Qy      301  ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db      901  CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Qy      321  GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db      961  CAGAGCCCTTGGGCAAGAGGTTACAGCTTAGACAGNAGCAACTTTTGAAGTGTGG 1020
Qy      341  GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db      1021  GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCAGGAGATTTGGT 1080
Qy      361  GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      1081  GGACCTCGCTCTTATACCATCGCACTTCTCCCTGGGTAAAGAGTGGCGCTGTAATTCCT 1140
Qy      381  AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db      1141  GCCTGCTTCATCACACAGCTCTCCTCGTGTGAAAGGAGCTAGAGTTCATGAATGAGCT 1200
Qy      401  SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db      1201  TCNAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCCTTCAGCTA 1251

RESULT 6
AK054547
LOCUS
DEFINITION Mus musculus 2 days pregnant adult female ovary cDNA, RIKEN
full-length enriched library, clone.E330039p08
product:galactosidase, alpha, full insert sequence.
ACCESSION AK054547
VERSION AK054547.1 GI:26344324
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
```

```
REFERENCE
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
PUBMED 10349636
REFERENCE
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
PUBMED 11042159
REFERENCE
AUTHORS Carninci,P., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Kono,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
PUBMED 11076861
REFERENCE
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,M.,
Hayaehida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,K., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
COMMENT
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/.
FEATURES
source
1..1376
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="C57BL/6J"
/db_xref="FANTOM_DB:E330039p08"
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.

```

/db_xref="taxon:10090"
/clone="E330039P08"
/sex="female"
/tissue_type="ovary"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="2 days pregnant adult"
21..1286
/note="unnamed protein product; galactosidase, alpha
(MGD|GI:1347344, GB|NM_013463, evidence: BLASTN, 99%,
match=1373)
putative"
/codon_start=1
/protein_id="BAC35819.1"
/db_xref="GI:26344325"
/translaton="MAMKLSRDLVCEALPLALFWSILGVRALDGLARTPTM
GWLHWFMCNLDQCEPDMACISQELFMQMAELMVSQWRDAGDYDILCIDDCWMAPE
DSKRLQADPORPPSGIKHLANYVHSKGLGIVADYGNKTCAGFGSGYDIDAOT
PADWGDLKPEDGCHCDVSVLENGKYMALALNRCTRSIVYSCWEPYLYLRPHKPNY
TDIYYCNHWRNFDVDVDSWESIKNLSWTVVYQKEIVVAGSGSNDPMLVIGNFG
LSWDQVQVQWALMAUAPLMLNDLRQISQAKALLQNDVIAINODPLGKQGYCFR
KENHIEYWERPLSNLAWAVRNLOETGGPCPYTIQISLGRGLACNPGCIITQLLPE
KVHLGFYEWTLLTKRVNPSGTVLFLER"
1358..1363
/note="putative"
1376
/note="putative"

```

CDS

polyA_signal

polyA_site

ORIGIN

```

Alignment Scores:
Pred. No.: 4,67e-187 Length: 1376
Score: 1806.00 Matches: 326
Percent Similarity: 88.01% Conservative: 41
Best Local Similarity: 78.18% Mismatches: 50
Query Match: 78.76% Indels: 0
DB: 4 Gaps: 0

```

US-10-602-220-16 (1-417) x AK054547 (1-1376)

```

Qy 1 MetGlnLeuArgAanProGluLeuHisLeuGlyCysAlaLeuAlaLeuAArgPheLeuAla 20
Db 27 ATGAAGCTTTTGACGAGATACCGGCTGGTCTGTGAGCTTGGCTTTGTCCTCGCT 86
Qy 21 LeuValSerTrpAspIleProGlyAlaAArgAlaLeuAspAenGlyLeuAlaAArgTrpPro 40
Db 87 TTAGTCTTCTGGAGCAATCTTGGGCTCAGACATTTGGACAATGGCTTGGCGGAGCTCT 146
Qy 41 ThrMetGlyTrpLeuHisTrpGluAArgPheMetCysAenLeuAaspCysGlnGluPro 60
Db 147 ACTATGGGCTGGCTGCAATTTGGAAACGTTTATGTGCAACCTTGACTGCCAAGAAGACCT 206
Qy 61 AspSerCysIleSerGluLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 207 GATGCTGCATAGTGAGCACTGTTCTATGAGATGCGAGCTCATGCTCTGTATGGC 266
Qy 81 TrpIleAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnAArg 100
Db 267 TGGCGGATGCGAGTTATGACTATCTCTGATAGATGACTGTTGGATGGCTCCCGAGAG 326
Qy 101 AspSerGluGlyAArgLeuGlnAlaAspProGlnAArgPheProHisGlyIleAArgGlnLeu 120
Db 327 GATTCAAGGCGAGCTTACGGCAGATCCCAACGCTTCTTCTAGTGGGATCAACACCTC 386
Qy 121 AlaAenTrpValHisSerIleGlyLeuLeuLeuGlyIleTrpAlaAspValGlyAenLys 140
Db 387 GCAATTTAGTCCACACCAAGGATTTGAGTAGGATTTATGCGAGATTTGGGAATAAA 446
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnTrpPheAla 160
Db 447 ACCTGTGAGGTTTCCCGGAGTTTGGATCTTATGACATTTGATGCGACATTTGCT 506
Qy 161 AspTrpGlyValAspLeuLeuLeuPheAspGlyCysTrpCysAspSerLeuGluAenLeu 180
Db 507 GACTGGGCGGTAGATCTGCTAAATTTGATGGTTGTCACTGTGACAGTGTAGTATCTCTG 566

```

```

Qy 181 AlaAspGlyTrpIleHisMetSerLeuAlaLeuAenAArgTrpGlyAArgSerIleValTrp 200
Db 567 GAGAATGTTTATAAGTACATGGCTTGGCTTGAACAGGACAGCCGCAAGCATTTGTATAC 626
Qy 201 SerCysGlnTrpProLeuTrpMetTrpProPheGlnLysProAenTrpTrpGluIleAArg 220
Db 627 TCCTGTGAGTGGCCACTTTATTTAGAGCCCTTTTATGAAGCCCAATATATACAGATATCAA 686
Qy 221 GlnTrpCysAenHisTrpAArgAenPheAlaAspIleAspSerTrpLysSerIleLys 240
Db 687 TATTACTGCAATCATTTGGAGAAATTTTATGATTTTATGATCTCTGGAAAGCATAAAG 746
Qy 241 SerIleLeuAspTrpTrpSerPheAsnGlnGluAArgIleValAspValAlaGlyProGly 260
Db 747 AATATCTTGTCTTGGACAGTGGTTTACAGAAAGAGATTTTGAAGTCGCTGGACCAAGC 806
Qy 261 GlyTrpAenAspProAspMetLeuValIleGlyAenPheGlyLeuSerTrpAenGlnGln 280
Db 807 AGCTGGAAATGCCAGACATTTAGTGTATGCGCACTTTGGCCTCAGTTGGGACCGAG 866
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 867 GTGACACAGATGGCCCTCTGGGCTATCATGCGGCTCCCTACTCATGTCCACAGATCTG 926
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAen 320
Db 927 CGACAAATCAGCTCTCAAGCCAAAGCTCTGCTTCAAGATTAAGGATGTAATTCCTCAAC 986
Qy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuAArgGlnGlyAspAenPheGluValTrp 340
Db 987 CAAGACCCCTTGGGCAAGAGGCTACTGTTTTCAGAAAGAAACCAACATTTGAGGTTGG 1046
Qy 341 GluAArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAenAArgGlnGluIleGly 360
Db 1047 GAAGGCCACTCTCCAACCTAGCTGGGCTGTGGCTGTGAGAAACCTGCAGGAGATTGGT 1106
Qy 361 GlyProAArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAenPro 380
Db 1107 GGACCTTGTCTTATACCATCCAGATTTCTTCCCTGGGTAGAGGACTAGCTGCAATCCT 1166
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysAArgLysLeuGlyPheTrpGluTrpThr 400
Db 1167 GGTGTCATCATTTACTAGCTTCTCCCGAGAAATGACACCTAGGCTTCTATGAATGACT 1226
Qy 401 SerAArgLeuAArgSerHisIleAenProThrGlyThrValLeuLeuGlnLeu 417
Db 1227 TTGACCTTAAAACTCGATTAACCTCTCAGGCACTGTTTGTTCGGTTA 1277

```

RESULT 7

AK040081

LOCUS

DEFINITION

sequence.

AK040081

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

TITLE

```

Genome Res. 10 (10), 1617-1630 (2000)
11042159
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kikunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
11076861
4 The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2962)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsumura, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, K., Takahashi, F., Takaku-Akachi, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/.
FEATURES
source
1. .2962
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:A430057F16"
/db_xref="taxon:10090"
/clone="A430057F16"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
1. .1266
/note="unnamed protein product; galactosidase, alpha
(MGI|MGI:1347344, GB|NM_013463, evidence: BLASTN, 99%,
match=1373)
putative"
/codon_start=1
/protein_id="BAC330508.1"
/db_xref="GI:26333581"
CDS
/translation="MAWKLLSRDLRVCELALCPALVFWLSILGVRALDNLGARTPTM
GWLHWFRCNLDCEPEPDACISEQFLMQMAELMVDGWRDAGDYLCIDDCWAPER
DSKGRLOADQPPRPSGIEKLANYVHSGKLGKLGIVADGNKTCAGFPSCFSGSYDIAOT
FADWGVDLLKFDSCDVSIVLENGKYMALALNRGTSIVYSCPEFLYLRPHKPNY
TDIQYCNWRNFDVYDSWESIKNLSVTYVYQKEIVEVAGSGWNPDMVLVGNFG
LSWDQQYQNALWAIWAPLLMSNDLRQISSQAKALLQNDKVDIAINQDPLCKGQCYFR
KENHIVRWPRLSNLAWAVARNLQIGGPGPYTIQISSIGRLGLACNPNCIIITOLLDE
KVHLGFYEWTLTLKTRVNPSTGLFLRLR"
ORIGIN
Alignment Scores:
Pred. No.: 1.5e-186 Length: 2962
Score: 1806.00 Matches: 326
Percent Similarity: 88.01% Conservative: 41
Best Local Similarity: 78.18% Mismatches: 50
Query Match: 0
DB: 4
US-10-602-220-16 (1-417) x AK040081 (1-2962)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 7 ATGAAGCTTTTGGAGCAGATACGGCGCTGCTGTGAGCTTGGCGCTTGTCCCTGGCT 66
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 67 TTAGTTTCTGGAGCAATCTTGGGGTCAGAGCATGGACAATGGCTTGGCGCGGACTCCT 126
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 127 ACTATGGCGCTGGCTGCAATGGGAACGTTTCATGTGCAACCTTGACTGCCAAGAGACGCT 186
QY 61 AspSerCysIleSerGluLeuPheMetGluAlaGluLeuMetValSerGluGly 80
Db 187 GATGCTTGCATAAGTAGGCAACTGTTTCATGCAATGCGAGAGCTCATGGTCTCTGATGGC 246
QY 81 TrpLeuAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 247 TGGCGGGATGACAGTTATGACTATCTGTCATAGATGACTGTTGGATGGCTCCGAGAGG 306
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 307 GATTCAAGGGCAGGCTTCAGGAGATGCCCAACGCTTCTCTAGTGGGATCAACACCTC 366
QY 121 AlaAsnTrpValHisSerLeuGlyLeuLeuGlyIleTrpAlaAspValGlyAsnLys 140
Db 367 GCAATTTAGCTCCACCAAGAGTTGAGCTAGGAGTTATGACAGATGTGGGAATAA 426
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
Db 427 ACCTGTGACAGTTTCCCGGGAGTTTGGATCCTATGACATTTGATGGCAGACATTGCT 486
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db 487 GACTGGGGCGTAGATCTGCTAAAAATTTGATGGTGTGCACCTGTGACAGTATGATCTTGT 546
QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
Db 547 GAGAAATGTTATAGTATACATGGCGCTTGGCCCTTGACAGGACAGCCGAGCATTTATAC 606
QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
Db 607 TCCTGTGAGTGGCCACTTTTATTTGAGACCCCTTTTCATTAAGCCCAATTTATACAGATATCAA 666
QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 667 TATTACTGCAATCATTTGAGAAATTTTGGATGATGTTTATGATCTCTGGGAAGCATTAAG 726
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGlyIleValAspValAlaGlyProGly 260
Db 727 AATATCTTGTCTGGACAGTGGTTTACCAGAGAGAGATTGTTGAAGTCGCTGGACAGGC 786
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280

```

```

Db      787 AGCTGGAATGACCCAGACATGTTAGTGATCGGCAACTTTGGCCCTCAGTTGGGACACGAG 846
Qy      281 ValThrGlnMetAlaLeuTrrAlaIleMetAlaAlaProLeuPheMetSerAspLeu 300
Db      847 GTGACACAGATGGCCCTCTGGGCTATCATGGCCGCTCCCTACTCATGTCCACAGATCTG 906
Qy      301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db      907 CGACAAATCAGCTCTCAAGCCAAAGCTCTGCTCAGAATAAGGATGTAATTCCTCAAC 966
Qy      321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrr 340
Db      967 CAAGACCCCTGGCAACGAGGCTACTGTTTCAGAAAGGAAACCACTTGAGGTTGG 1026
Qy      341 GluArgProLeuSerGlyLeuAlaTrrAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db      1027 GAAGGCCACTCTCCACCTAGCTGGGCTGTGGCTGTGAGAAACCTGCAGGAGATGGT 1086
Qy      361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      1087 GGACCTTGCTTATACCATCCAGATTTCTTCCCTGGGTAGAGGACTAGCCTGCAATCCT 1146
Qy      381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrrThr 400
Db      1147 GGCTGTCATCATTTACTAGCTTCTCCCGAGAAAGTACACCTAGGCTTCTATGAATGACT 1206
Qy      401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db      1207 TTGACCTTAAAACTCAGTAAACCCCTCAGGCACTGTTTGTTCGGTTA 1257

```

RESULT 8

AL554978

LOCUS

AL554978 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
 CNA clone CS0DK009YE12 5-PRIME, mRNA sequence.

ACCESSION

AL554978

VERSION

AL554978.3

KEYWORDS

EST.

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

REFERENCE

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

On Feb 15, 2001 this sequence version replaced gi:31276787.

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

10506.r

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?rs=CS0DK009BC06QPl&c=10506.r.

FEATURES

Location/Qualifiers

1..1086

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clones="CS0DK009YE12"

/cell_type="HELA CELLS COT 25-NORMALIZED"

/cell_line="HELA"

/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and EcoR V

digested with Not I and EcoR V

digested with Not I and EcoR V

digested with Not I and EcoR V

digested with Not I and EcoR V

digested with Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

Pred. No.: 5,02e-184 Length: 1086

Score: 1777.00 Matches: 342

Percent Similarity: 96.65% Conservativity: 4

Best Local Similarity: 95.53% Mismatches: 10

Query Match: 77.50% Indels: 7

DB: 1 Gaps: 1

US-10-602-220-16 (1-417) x AL554978 (1-1086)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20

Db 6 ATGCAGCTGAGGAACCCAGAACATACATCTGGGCTGCGCGCTTGC-CTTCGCTTCTCGGCC 64

Qy 21 LeuValSerTrrPheIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40

Db 65 CTCGTTTCTCGGACATCCCTCGGGCTAGAGCACTGGCAATGATGATGGCAAGGACGCT 124

Qy 41 ThrMetGlyTrrLeuHisTrrGluArgPheMetCysAsnLeuAspCysGlnGluPro 60

Db 125 ACCATGGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCA 184

Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80

Db 185 GATTCTCTGCATCAGTGAGAAGCTTTCATGAGAGATGGCAGAGCTCATGGTCTCAGAAGGC 244

Qy 81 TrrPheAspAlaGlyTrrGluTrrLeuCysIleAspAspCysTrrMetAlaProGlnArg 100

Db 245 TGGAAGGATCAGGATTATGATGACCTCTGCAATGATGACTGTGGATGGCTCCCAAGA 304

Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120

Db 305 GATTCAAGGACGACATTCAGGAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 364

Qy 121 AlaAsnTrrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140

Db 365 GCTAAATTATGTTACACAAAGGACTGAAGTAGGATTTATGACAGATGTTGGAAATATA 424

Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrrTrrAspIleAspAlaGlnThrPheAla 160

Db 425 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTTGATGACAGCTTTGG 484

Qy 161 AspTrrGlyValAspLeuLeuLysPheAspGlyCysTrrCysAspSerLeuGluAsnLeu 180

Db 485 GACTGGGAGTAGATCTGCTTAAATTTGATGTTGTTACTGTGACAGCTTTGGAAATTTG 544

Qy 181 AlaAspGlyTrrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200

Db 545 GCAGATGGTTATAAGCACATGTCCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 604

Qy 201 SerCysGluTrrProLeuTrrMetTrrProPheGlnLysProAsnTrrThrGluIleArg 220

Db 605 TCCTGTGAGTGGGCTCTTTATATGTGGCCCTTTCAAAGGCCCAATATACAGAATTCGA 664

Qy 221 GlnTrrCysAsnHisTrrArgAsnPheAlaAspIleAspAspSerTrrLysSerIleLys 240

Db 665 CAGTACTGCATCTACTGGCGAATTTTGTGCTGACATTCATGATTCCTGGAAAGTATAAG 724

Qy 241 SerIleLeuAspTrrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260

Db 725 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTGATGTTGTCTGGACCGG 784

Qy 261 GlyTrrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrrAsnGlnGln 280

Db 785 GGTGGAAATCACCAGATATGTTAGTATGGCAACTTTTGGCCCTCAGCTGGAATCAGCAA 844

Qy 281 ValThrGlnMetAlaLeuTrrAlaIle-MetAlaAlaProLeuPheMetSerAsnAspLe 300

Db 845 GTAACTCAGATGGCCCTCTGGGSTATTTATCATGGCTGCTCTTCTTATTCATGCTAATGACT 904


```
/clone lib="NIH_MGC_119"
/notes="Organ: brain; Vector: pCMV-SPORT6; Site1: NotI;
Site2: EcoRV (destroyed); RNA source normal medulla from
anonymous male age 27. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.3 kb, insert size range
0.9-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 013. Note:
this is a NIH MGC Library."
```

ORIGIN

Alignment Scores:		
Pred. No.:	5,318-172	Length:
Score:	1667.50	Matches:
Percent Similarity:	93.86%	Conservative:
Best Local Similarity:	91.52%	Mismatches:
Query Match:	72.72%	Indels:
DB:	3	Gaps:
		0
		1067
		313
		8
		19
		4
		0

US-10-602-220-16 (1-417) x BM564282 (1-1067)

Qy	3	LeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAlaLeuVal	22
Db	16	CTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGCCTTCCTGGGCCCTCGTT	75
Qy	23	SerTrpAspIleProGluValaargAlaLeuAspAsnGlyLeuAlaArgThrProThrMet	42
Db	76	TCTTGGGACATCCCTGGGGCTAGAGCACTGGACAAATGSAATTGGCAGAGCGCTTACCATG	135
Qy	43	GlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluProAspSer	62
Db	136	GGCTGGCTGCATCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGACCCAGATTCC	195
Qy	63	CysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLys	82
Db	196	TGCATCAGTGAGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGCTCGAAG	255
Qy	83	AspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArgAspSer	102
Db	256	GATGCAAGGTATGAGTACCTCTGCATTGATGACTGTTGGATGGCTCCCCAAAGAGATTCA	315
Qy	103	GluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeuAlaAsn	122
Db	316	GAAGGCAGACTTCAGGCACCCCTCAGCGCTTCTCATGGGATTGCCACAGCTAGCTAAT	375
Qy	123	TyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLysThrCys	142
Db	376	TATGTTTCACAGCAAAAGGACTGAAGCTAGGGATTTTATGCAGATGTTGGAAATAAAACCTGC	435
Qy	143	AlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAlaAspTrp	162
Db	436	GCAGGCTCCCTCGGAGTTTTGGATPACTACACAACTGATGCCACAGCTTTGTCTGACTGG	495
Qy	163	GlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeuAlaAsp	182
Db	496	GGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGCACAGTTTGGAAAAATTTGGCAGAT	555
Qy	183	GlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyrSerCys	202
Db	556	GGTTATAAGCACATGTCCTTGCCCTGGCCCTGAATAGGACTGGCAGAAGCATTTGTGTACTCCTGT	615
Qy	203	GluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArgGlnTyr	222
Db	616	GAGTGGCCCTTTATATGTGGGCCCTTTTCAAAAGGCCCAATTATACAGAAAAATCCGACAGTAC	675
Qy	223	CysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLysSerIle	242
Db	676	TGCATACACTGGCGAAATTTTCTGCATCATGATGATCTCGAAAGAGTATTAAGAGTATC	735
Qy	243	LeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGlyGlyTrp	262
Db	736	TTTGGACTGCACATCTTTTAAACAG-CAGAGAAATTTGTGATGTTGCTGCAGCCAGGGGGGTG	794


```
Best Local Similarity: 95.71% Mismatches: 5
Query Match: 72.37% Indels: 6
DB: 3 Gaps: 1

US-10-602-220-16 (1-417) x BQ062192 (1-1005)

Qy 28 GlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrProThrMetGlyTrpLeuHisTrp 47
Db 1 GGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCTACCAATGGCTGGCTGGCTGCACTGG 60
Qy 48 GluArgPheMetCysAsnLeuAspCysGlnGluLeuProAspSerCysIleSerGluLys 67
Db 61 GAGCGCTTCATGTGCAACCTTGACTGCCAGAAAGACGAGATCTCTCATCAGTGAGAG 120
Qy 68 LeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLysAspAlaGlyTrpGlu 87
Db 121 CTCTTCATGAGATGGCAGAGCTCATGCTCAGAAGGCTGGAAGGATGCGAGGTATGAG 180
Qy 88 TyrLeuCysIleAspAspCysTrpMetAlaProGlnArgAspSerGluGlyArgLeuGln 107
Db 181 TACCTCTGCATTGATGACTGTGGATGGCTCCCAAGAGATTTCAGAAAGGACAGACTTCAG 240
Qy 108 AlaAspProGlnArgPheProHisGlyLeuArgGlnLeuAlaAsnTrpValHisSerLys 127
Db 241 GCAGACCTCAGCGCTTCTCTCATGGATTCGCCAGCTAGCTAATTAATGTTTCAGCAAA 300
Qy 128 GlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLysThrCysAlaGlyPheProGly 147
Db 301 GGACTGAGCTAGGATTTATGAGATGTTGGAATTAACCTGGCGAGGCTTCCCTGGG 360
Qy 148 SerPheGlyTyrTrpAspIleAspAlaGlnThrPheAlaAspTrpGlyValAspLeuLeu 167
Db 361 AGTTTGGATACACGACATTGATGCCAGACCTTCTGCTGCTGGGAGTAGATCTGCTA 420
Qy 168 LysPheAspGlyCysTrpCysAspSerLeuGlnAsnLeuAlaAspGlyTyrLysHisMet 187
Db 421 AAATTTGATGGTTGTTACTGTGACAGTTGGAATTTGGCAGATGGTTAATAGCACATG 480
Qy 188 SerLeuAlaLeuAsnArgThrGlyArgSerIleValTrpSerCysGluTrpProLeuTyr 207
Db 481 TCCTTGGCCCTGATAGACATGGCAGAGCAATGTGTACTCTCTGTGATGGCTCTTTAT 540
Qy 208 MetTrpProPheGlnLysProAsnTrpThrGluIleArgGlnTrpCysAsnHisTrpArg 227
Db 541 ATGTGGCCCTTCAAAGGCCCAATTATACAGAAATCCGACAGTACTGCAATCACTGGCGA 600
Qy 228 AsnPheAlaAspIleAspAspSerTrpLysSerIleLysSerIleLeuAspTrpThrSer 247
Db 601 AATTTTGTCTGACATTGATGATTCCTGGAAAGTATAAAGAGTATCTTTGGACTGGACATCT 660
Qy 248 PheAsnGlnGluArgIleValAspValAlaGlyProGlyGlyTrpAsnAspProAspMet 267
Db 661 TTTAACCCAGGAGAGAAATTTGATGTTGCTGGACAGGGGTTGGAATGATGATGATGATG 720
Qy 268 LeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnValThrGlnMetAlaLeuTrp 287
Db 721 TTAGTGATGGCAACTTGGCCCTCAGCTGGAATCAGCAAGTAACCTCAGATGGCCCTCTGG 780
Qy 288 AlaIleMetAlaAlaProLeuPheMetSerAsnAspLeuArgHisIleSerProGlnAla 307
Db 781 GCTATATGCTGCTCTCTTTATCATGTCTAATGACCTCCGACACATCAGCCCTCAAGCC 840
Qy 308 LysAlaLeuLeuGlnAspLysAspValIleAla-IleAsnGlnAspProLeu-GlyLysG 327
Db 841 AAAGCTCTTCTCAGGATTAAGAGCTTATTTGCCCATCAATCAGGACCCCTTTGGGCAAGC 900
Qy 327 In-GlyTyrGlnLeu-ArgGlnGly-AspAsnGlyLeu-ValTrpGluArgProLeuS 345
Db 901 AAAGTTTACCAGCTTTAAACAGGGGAAACCACTTTGAAGGGTGGGAAACGACCCCTCTCT 960
Qy 345 erGlyLeu 347
Db 961 CAGGGCTA 968
```

CO645623 922 bp mRNA linear EST 23-JUL-2004
ILLUMIGEN MQ0.24642 Katze MMLV Macaca mulatta cDNA clone
TBIUM:23923 5' similar to Bases 5 to 920 highly similar to human
GLA (Hs.69089), mRNA sequence.

CO645623 1 GI:50567117
EST.
Macaca mulatta (rhesus monkey)
Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Cercopithecoidea; Cercopithecinae; Macaca.
1 (bases 1 to 922)
Magnes, C.L., Fellin, P.C., Thomas, M.J., Korth, M.J., Agy, M.B.,
Proli, S.C., Fitzgibbon, M., Scherer, C.A., Miner, D.G., Katze, M.G. and
Tadonato, S.P.
Analysis of the Macaca mulatta transcriptome and the sequence
divergence between Macaca and human
Genome Biol. 6 (7), R60 (2005)
15998449
Contact: C. Magnes
Illumigen Biosciences Inc.
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
Tel: 2063780400
Fax: 2063780408
Email: cmagnes@illumigen.com
Sequenced on 2004.03.02. 796 Q20 bases. Library Preparation: Prof.
Michael Katze lab at University of Washington DNA Sequencing:
Illumigen Biosciences Inc. For further information, see
http://www.macaque.org
PCR Primers
FORWARD: CCTCTACTAAAGGGACAAAA
BACKWARD: CACTATAGGCGAATTGGTA
Insert length: 922 Std Error: 0.00
Plate: CL000111 row: E column: 08
Seq primer: CCTCTACTAAAGGGACAAAA
POLYA=No.

Location/Qualifiers
1. 922
/organism="Macaca mulatta"
/mol_type="mRNA"
/strain="Indian"
/db_xref="taxon:9544"
/clone="TBIUM:23923"
/sex="female"
/dev_stage="adult"
/lab_host="Electromax DH10B"
/clone_lib="Katze MMLV"
/notes="Organ: liver; Vector: pDONR 222; Site:1: BarG I;
Site 2: BarG I; Created from CloneMiner cDNA Library
Construction kit (catalog #18249-029)"

Alignment Scores:
Pred. No.: 4,59e-170 Length: 922
Score: 1649.00 Matches: 295
Percent Similarity: 98.03% Conservative: 3
Best Local Similarity: 97.04% Mismatches: 6
Query Match: 71.91% Indels: 0
DB: 7 Gaps: 0

US-10-602-220-16 (1-417) x CO645623 (1-922)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 11 ATGCAGCTGAGGACCCAGAAATGCATCTGGCTGGCGCTTCTCTTCTCTGGCC 70
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 71 CTCGTTTCTGGGACATCTCTGGGGCCAGACACTGGCAATGGATTGGCAGACGCT 130

FEATURES
source

ORIGIN

Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 Db 131 ACCATGGGCTGGCTGGCACTGGAGCGCTTCATGTGCAACCTTGACTCCAGAGAGACCA 190
 Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db 191 GATTCCTGCATCAGTGAGAGCTCTTCATGAGATGCGAGAGCTCATGGTCTCAGATGGC 250
 Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 251 TGGAAAGGATCAGGTTATGATGATCTCTGCAATGATGACTGTGTGGATGGCTCCCCAAAGA 310
 Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 Db 311 GATTTAGAGGCAGACTTCAGGCGAGACCCCTCAGCGCTTTCTCATGGGATTCGCCAGCTA 370
 Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 Db 371 GCTAATATGTTACACCAAGGAGCTGAGCTAGGATTTATGCAGATGTTGGAATATAA 430
 Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
 Db 431 ACCTGCCAGGCTTCCCTGGGAGTTTGGATCTATGACATTTGATGCCAGACCTTTGCT 490
 Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 Db 491 GACTGGGAGTAGATCTGCTGAAATTTGATGGTTTGTACTGTGACAGTTTGGAAAGTTG 550
 Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 Db 551 GCAGATGTTATAGCACATGTCCTTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC 610
 Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
 Db 611 TCCTGTGAGTGGCTCTTTATATGTGGCCATTTCAAAGGCCCAATTACACAGAAATCCGA 670
 Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 Db 671 CAGTACTGCATCACTCGCGAATTTTGTGCATGATGATCTCTGGAAGATTAAG 730
 Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 Db 731 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTGATGTTGTGGACAGGG 790
 Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 Db 791 GGTTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 850
 Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 Db 851 GTRACTCAGATGGGCCCTCTGGCTATCATGCTCTCTTTATTCATGCTTAATGACCTC 910
 Qy 301 ArgHisIleSer 304
 Db 911 CGACACATCACC 922

RESULT 14
 AL577581/c 1026 bp mRNA linear EST 07-APR-2004
 LOCUS AL577581 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
 DEFINITION CDNA clone CS0DK009VE12 3-PRIME, mRNA sequence.
 ACCESSION AL577581.3 GI:46256603
 VERSION AL577581.3
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Hominidae; Homo.
 1 (bases 1 to 1026)
 Li,W.B., Gruber,C., Jesses,J. and Polayes,D.
 Full-length CDNA libraries and normalization

JOURNAL COMMENT

Unpublished (2001)
 On Feb 16, 2001 this sequence version replaced gi:31315840.
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 2 rue Gaston Crémieux, CP 5706 - 91057 EVRY cedex - FRANCE
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen. This sequence belongs to sequence cluster
 10506.r

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?s=CS0DK009BC06NFI&c=10506.r.

FEATURES

source

1..1026
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DK009VE12"
 /cell_type="HELA"
 /cell_line="HELA"
 /clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
 /notes="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:
 Pred. No.: 6..14e-170 Length: 1026
 Score: 1648.50 Matches: 308
 Percent Similarity: 98.41% Conservative: 1
 Best Local Similarity: 98.09% Mismatches: 5
 Query Match: 71.89% Indels: 2
 DB: 1 Gaps: 0

US-10-602-220-16 (1-417) x AL577581 (1-1026)

Qy 97 AlaProGlnArgAspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGly 116
 Db 955 GCTCCCCAAAGAGATTGAGAACG-AGACTTCAGCA-GACCTTCAGCCCTTTCCTCATGGG 898
 Qy 117 IleArgGlnLeuAlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAsp 136
 Db 897 ATTGCGCAGCTAGCTAATATGTTACAGCAAGGACTGAAGCTAGGATTTATGACAGAT 838
 Qy 137 ValGlyAsnLysThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAla 156
 Db 837 GTTGGAAATAAACCTCGCAGCGCTTCCCTGGGAGTTTGGATACTACGACATTTGATGCC 778
 Qy 157 GlnThrPheAlaAspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSer 176
 Db 777 CAGACCTTTGCTGACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGT 718
 Qy 177 LeuGluAsnLeuAlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArg 196
 Db 717 TTGGAAATTTGGCAGATGTTATAGCACATGCTCTTGGCCCTGAATAGACTGGCAGA 658
 Qy 197 SerIleValTyrSerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyr 216
 Db 657 AGCATTTGTACTCTCTGAGTGGCCCTCTTTATATGTGGCCCTTTTCAAAAGCCCAATTAT 598
 Qy 217 ThrGluIleArgGlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrp 236
 Db 597 ACAGAAATCCGACAGTACTGCAATCTGCGGAATTTTGTGTCATGATGATGATCTCTGG 538
 Qy 237 LysSerIleLysSerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspVal 256
 Db 537 ABAAGTATAAGAGTATCTTGGACTGGACATCTTTTAAACAGAGAGAGAAATTTGTTGATGTT 478
 Qy 257 AlaGlyProGlyGlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSer 276

```

Db 477 GCTGACACAGGGGTTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCTCAGC 418
Qy 277 TTPAenGlnValThrGlnMetAlaLeuTriAlaIleMetAlaAlaProLeuPheMet 296
Db 417 TGGAAATCAGCAAGTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTTTATTTCATG 358
Qy 297 SerAenAspLeuArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspVal 316
Db 357 TCTAATGACCTCCGACACATCAGCCCTCAAGCCAAAGCTCTCTTACAGGATAAGGAGCTA 298
Qy 317 IleAlaIleAenGlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAen 336
Db 297 ATTGCCATCAATCAGGACCCCTTGGCAAGCAAGGGTACCAGCTTAGACAGGAGAGCAAC 238
Qy 337 PheGluValTriPLeuArgProLeuSerGlyLeuAlaTriPLeuAlaValAlaMetIleAenArg 356
Db 237 TTTGAAGTGTGGAAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGG 178
Qy 357 GlnGluIleGlyGlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyVal 376
Db 177 CAGGAGATTGGTGCACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTG 118
Qy 377 AlaCysAenProAlaCysPheIleThrGlnLeuLeuProValLysArgGlyLysLeuGlyPhe 396
Db 117 GCCTGTAAATCTGCTGCTCTCATCACACAGCTCTCMTGTGAAAGGAAGCTAGGGTTC 58
Qy 397 TyrGluTriPThrSerArgLeuArgSerHisIleAenProThr 410
Db 57 TATGATGGACTTCAGGTTAAGNAGTAACTAATAATVCRACA 16

RESULT 15
AY408542 1296 bp DNA linear GSS 15-DEC-2003
LOCUS Mus musculus GLA gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY408542
VERSION AY408542.1 GI:39764513
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Murinae; Mus.
1 (bases 1 to 1296)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
    source
        1..1296
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            <1..>1296
            /gene="GLA"
            /locus_tag="HMC3258"
gene
ORIGIN
Alignment Scores:

```

```

Pred. No.: 2,02e-167 Length: 1296
Score: 1627.00 Matches: 300
Percent Similarity: 81.06% Conservative: 38
Best Local Similarity: 71.94% Mismatches: 79
Query Match: 70.96% Indels: 0
DB: 10 Gaps: 0

US-10-602-220-16 (1-417) x AY408542 (1-1296)
Qy 1 MetGlnLeuArgAenProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGAAGCTTTTGACGACAGATACGCGCTGGTCTGTGAGCTTGGCTTTGGCTTGGCT 60
Qy 21 LeuValSerTriPAspIleProGlyAlaArgAlaLeuAspAenGlyLeuAlaArgThrPro 40
Db 61 TTAGTTTCTGGAGCATCTCTGGGGTTCAGAGCATTTGACAAATGGCTTGGCGGACTCCT 120
Qy 41 ThrMetGlyTriPLeuHisTriPLeuArgPheMetCysAenLeuAspCysGlnGluPro 60
Db 121 ACTATGGGCTGGCTGCTATTTGGGAACGTTTCATGTGCAACCTTGACTCCCAAGAGAGCCT 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATGCTGCATAAGTGGCAACTGTTTCATGCAGATGGCAGAGCTCATGGTCTCTGATGGC 240
Qy 81 TriPLeuAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTriPMetAlaProGlnArg 100
Db 241 TGGCGGAGTGCAGGTTATGACTATCTGTCATAGATGACTGTTGGATGGCTCCCGAGAGG 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAAGGGGAGGCTTCAGGAGATCCCAACGCTTCTCTAGTGGGATCAACACCTC 360
Qy 121 AlaAenTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAenLys 140
Db 361 GCAATTCGTCACAGCAAGGATTTGAAGTAGGATTTATGCAGATGTTGGGAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGTGAGGTTTCCCGGAGTTTGGATCTATGACATTTGATGGCAGACATTTGCT 480
Qy 161 AspTriPLeuValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAenLeu 180
Db 481 GACTGGGGCGTAGATCTGCTAAATTTGATGGTGTGCTGACAGTGTAGTATCTCTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAenArgThrGlyArgSerIleValTyr 200
Db 541 GAGAAATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 600
Qy 201 SerCysGluTriPProLeuTyrMetTriPProPheGlnLysProAenTyrThrGluIleArg 220
Db 601 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 660
Qy 221 GlnTyrCysAenHisTriPArgAenPheAlaAspIleAspAspSerTriPLeuSerIleLys 240
Db 661 TATTACTGCAATCATTTGGGAAATTTTGAATGATTTATGATCTCTCGGGAAGCATAAAG 720
Qy 241 SerIleLeuAspTriPThrSerPheAenGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AATATCTTGTCTGGACAGTGGTTTACCAGAGAGATTTGTTGAAGTGGCTGGACAGGC 780
Qy 261 GlyTriPAsnAspProAspMetLeuValIleGlyAenPheGlyLeuSerTriPAsnGlnGln 280
Db 781 AGCTGGAAATGACCCAGACATGTTAGTGTGCGCAACTTTGGCCTCAGTTGGACAGCAG 840
Qy 281 ValThrGlnMetAlaLeuTriPAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTGACACAGATGGCCCTCTGGGCTATCATGCGCGCTCTCTCTACTCTCATGTCTCAACGATCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAen 320
Db 901 CGAATAATCAGCTCTCAAGCCAAAGCTCTCTTCAAGATTAAGGATGTAATTTGCCATCAAC 960

```

```
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
961 CAAGACCCCTTGGCAAGCAGGGCTACTGTTTCAGAAAGGAAACACCATTTGAGTTTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTTPAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1021 GNACGGCCACTCTCCAACTAGCCTGGGCTGGGCTGTGAGAAACCTGCAGGAGATTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1081 GGACCTTTGTCCTTATACCATCCAGATTCTTCCCTGGGTAGAGGACTAGCCTGCAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1141 GGCTGCATCATTAATCAGCTTCTCCCGAGAAAGTACACCTAGGCTTCTATGATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1201 TTGACCTTAAAAAACTCGAGTAACCCCTCAGGCACCTGTTTGTTCGGTTA 1251
```

Search completed: January 1, 2006, 05:47:02
Job time : 7590.48 secs

PA (GARG/) GARGER S J.
 PA (TURP/) TURPEN T H.
 PA (KUMAI/) KUMAGAI M H.
 XX
 PI Garger SJ, Turpen TH, Kumagai MH;
 DR WPI: 2002-681656/73.
 DR P-PSDB: A828212.
 XX
 PT Novel human alpha-galactosidase polypeptide useful for treating lysosomal
 PT storage diseases.
 XX
 PS Claim 1; Page 47-49; 88pp; English.
 XX
 CC The invention relates to human alpha-galactosidase truncated at the
 CC carboxy terminus and the production of enzymatically active recombinant
 CC human and animal lysosomal enzymes. The invention is useful for producing
 CC lysosomal enzymes for treating lysosomal storage diseases, producing
 CC altered or mutated proteins, enzymatically active or otherwise, to serve
 CC as precursors or substrates for further in vivo or in vitro processing to
 CC a specialised industrial form for research or therapeutic uses, to
 CC produce more effective therapeutic enzyme, for producing antibodies
 CC against lysosomal enzymes for medical diagnostic use, and in any
 CC commercial process that involves substrate hydrolysis. The present
 CC sequence is human rGAL-12 DNA
 XX
 SQ Sequence 1254 BP; 316 A; 289 C; 328 G; 321 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2,346-235 Length: 1254
 Score: 2293.00 Matches: 417
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0

US-10-602-220-16 (1-417) x A4D5225 (1-1254)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTCGGCC 60
 QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 Db 61 CTGCTTCTCTGGGACATCCCTGGGGCTAGAGCACTGGACATGATGGTGGCAAGAGCCCT 120
 QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 Db 121 ACCATGGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
 QY 61 AspSerCysIleSerGluLeuLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db 181 GATTCTGCATCAGTCAGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
 QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 241 TGAAGAGATGCAGGTTATGATGATACCTCTGATTTGATGACTTTGGATGGCTCCCAAGA 300
 QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 Db 301 GATTCAAGAGCAGACTTCAGGCAGACCCCTCAGCGCTTTCTCATGGATTGCCAGCTA 360
 QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 Db 361 GCTAAATATGTTACACAGCAAGGACTGAAGCTAGGGATTTATGCAGATGTGGAAATAAA 420
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
 Db 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATATACGACATTTGATGCCAGACTTTGCT 480
 QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 Db 481 GACTGGGAGTAGATCTGCTAAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAAATTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 Db 541 GCAGATGGTTATAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGAGCACTTGTGTAC 600
 QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
 Db 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
 QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 Db 661 CAGTACTGCATCAGTGGCGAAATTTGTCTGACATTTGATGATTTCTCGGAAAGTATAAAG 720
 QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 Db 721 AGTATCTTGGATCGACATCTTTTAACAGAGAGAGAAATTTGTATGTTCTGGACACGGG 780
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 Db 781 GGTGGAAATGACCCAGATATGTTAGTATTTGGCACTTTGGCCCTCAGCTGGAATCAGCAA 840
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCATAATGACCTC 900
 QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGCTAATTTGCCATCAAT 960
 QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACCACTAGACAGGAGACAACTTTGAAGTGTGG 1020
 QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGATGATTAACCGGAGGAGATTGGT 1080
 QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTGGGTAAAGGAGTGGCTGTATCTCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
 Db 1141 GCCTGCTTCATCACAGACTCTCTCCCTGTCAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
 Db 1201 TCAAGGTTAAGAGTCACATAATCCACAGGCACTGTTTTCCTTCAGCTA 1251

RESULT 2
 ADD84754
 ID ADD84754 standard; DNA; 1254 BP.
 XX
 AC ADD84754;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Human alpha-galactosidase rGAL-12 DNA.
 XX
 KW Human; alpha-galactosidase; rGAL-4; gene; ds; lysosomal enzyme;
 KW enzyme replacement therapy; lysosomal disease.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1254
 FT /*tag= a
 FT /product= "Human rGAL-12"
 XX
 PN US2003106095-A1.
 XX
 PD 05-JUN-2003.
 XX

PF 20-MAR-2002; 2002US-00103327.
 XX
 PR 26-JUL-2000; 2000US-00626127.
 PR 13-NOV-2001; 2001US-00993059.
 XX
 PA (GARG/) GARGER S J.
 PA (TURP/) TURPEN T H.
 PA (KOMA/) KUMAGAI M H.
 XX
 XX Garger SJ, Turpen TH, Kumagai MH;
 PI
 XX WPI; 2003-801257/75.
 DR P-PSDB; ADD84755.
 DR
 XX
 PT New polynucleotide for producing active recombinant human and animal
 PT lysosomal enzymes in a plant expression system that can be used in enzyme
 PT replacement therapy.
 XX
 XX
 XX Claim 1; SEQ ID NO 15; 77pp; English.
 XX
 CC The invention relates to human alpha-galactosidase derivatives and the
 CC nucleic acids encoding them. The polypeptides are used in a method for
 CC producing active recombinant human and animal lysosomal enzymes in a
 CC plant expression system. The enzymes can be used in enzyme replacement
 CC therapy for the therapeutic treatment of human and animal lysosomal
 CC diseases. This sequence represents DNA encoding a human alpha-
 CC galactosidase derivative polypeptide of the invention.
 XX
 SQ Sequence 1254 BP; 316 A; 289 C; 328 G; 321 T; 0 U; 0 Other;
 SQ

Alignment Scores:
 Pred. No.: 2,34e-235 Length: 1254
 Score: 2293.00 Matches: 417
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-602-220-16 (1-417) x ADD84754 (1-1254)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTCCGCTTCCCTGGCC 60
 QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 DB 61 CTGCTTCTCGGACATCCCTGGGGCTAGAGCACTGGACATGATGGTGGCAAGAGCCCT 120
 QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 DB 121 ACCATGGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
 QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 DB 181 GATTCTGTCATCAGTGAAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
 QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 DB 241 TGGAGGATGTCAGGTATGATGATCTTCTGATGATGATGATGATGATGATGATGATGAT 300
 QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 DB 301 GATTCAGAGGAGAGCTTCAGGCAGACCCCTCAGCGGCTTCTCATGGGATTCGCCAGCTA 360
 QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 DB 361 GCTAATATGTTTCACAGCAAGAGCTGAGCTAGGATTTATGACATGTTGGAAATAAA 420
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
 DB 421 ACCTGGCAGGCTTCCCTGGAGTTTGGATATACGACATGATGATGCCAGACCTTGGT 480
 QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180

DB 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTACAGTATTGGAAATTTG 540
 QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 DB 541 GCAGATGGTTATTAAGCACATGCTCTGGCCCTGAATAGGACTGGCAGAGCATTTGTATC 600
 QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
 DB 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTCAAAAGCCCAATTATACAGAAATCCGA 660
 QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 DB 661 CAGTACTGCAATCACCTGGCGAAATTTGTGTGACATGATGATTCCTGGAAAGATATAAG 720
 QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 DB 721 AGTATCTTGGACTGGACATCTTTTAACAGGAGAGAAATTTGTGATGTTGTGGACAGG 780
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 DB 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCACTTTGGCCCTCAGCTGGAATCAGCAA 840
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
 QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 DB 901 CGACACATCAGCCCTCAGCCCAAGCTCTCTTCAGGATAAGGACGTAATTCATCATCAT 960
 QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 DB 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
 QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 DB 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGCTGTAGCTATGATAAACCCGCGAGAGTTGGT 1080
 QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 DB 1081 GGACCTCGCTCTTATACCATCGCATGTTGCTTCCTGGGTAAAGAGTGGCTGTATCTCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
 DB 1141 GCCTGCTTCATCACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
 DB 1201 TCAAGGTTAAGAAGTCACATNAATCCACAGGCACTGTTTGTCTTCAGCTA 1251

RESULT 3
 ID ADM48684 standard; DNA; 1254 BP.
 XX ADM48684;
 XX 03-JUN-2004 (first entry)
 XX Human wild type rGAL-12 DNA.
 XX
 KW Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
 KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
 KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
 KW gene; ds.
 XX Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FT 1. .1254
 FT /*tag= a
 FT /product= "GAL-12 protein"
 XX

XX 10-FEB-2005 (first entry)
 XX DE Human alpha-galactosidase protein encoding DNA #7.
 XX
 XX Lysosomal enzyme; glucocerebrosidase; GCB; GCB; alpha-galactosidase;
 KW enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;
 KW Niemann-Pick disease; Fabry's disease; Tay-Sachs disease;
 KW Hurler's syndrome; Hurler-Scheie syndrome; nephrotropic; human; gene; ds.
 XX OS Homo sapiens.
 XX

Key Location/Qualifiers
 CDS 1..1254
 FT /*tag= a
 FT /product= "Human alpha-galactosidase protein"

US2004234516-A1.

25-NOV-2004.

21-MAY-2004; 2004US-00851388.

26-JUL-2000; 2000US-00626127.

13-NOV-2001; 2001US-00993059.

20-MAR-2002; 2002US-00103327.

(LARG-) LARGE SCALE BIOLOGY CORP.

Garger SJ, Turpen TH, Kumagai MH;

WPI; 2004-821274/81.

P-PSDB; ADU66920.

A pharmaceutical composition comprising a lysosomal enzyme, useful for
 PT enzyme replacement therapy for the treatment of lysosomal storage
 PT diseases, such as Fabry's disease.

PS Disclosure; SEQ ID NO 15; 88pp; English.

CC The present invention relates to the production of human and animal
 CC lysosomal enzymes in plants by a transient plant expression system. The
 CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
 CC galactosidase enzymes having a post-translational modification provided
 CC by the plant expression system. The invention is useful in enzyme
 CC replacement therapy for treating lysosomal storage diseases such as
 CC Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs
 CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
 CC also useful in researches for developing new approaches to medical
 CC treatment of lysosomal storage diseases and in industrial processes
 CC involving enzymatic substrate hydrolysis. The present sequence is the
 CC human alpha-galactosidase protein encoding DNA.

SQ Sequence 1254 BP; 316 A; 289 C; 328 G; 321 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2..34e-235	Length:	1254
Score:	2293.00	Matches:	417
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	13	Gaps:	0

US-10-602-220-16 (1-417) x ADU66919 (1-1254)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTCGGCC 60
 QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 DB 61 CTCGTTCTTCGGGACATCCCTGGGGCTAGAGCACTGGACAACTGGATTGGCAAGAGCGCCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 DB 121 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTCAACCTTGACTGCCAGAGAGCCA 180
 QY 61 AspSerCysIleSerGluLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 DB 181 GATTCTGTCATCATGTAGAGAGCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
 QY 81 TrpLeuAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 DB 241 TGAAGGATGACGGTTATGAGTACCTCTGTCATTGATGACTGTTGGATGGCTCCCAAGA 300
 QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 DB 301 GATTCAAGAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCAGCTA 360
 QY 121 AlaAsnTrpValHisSerIleGlyLeuGlyLeuGlyIleTrpAlaAspValGlyAsnLys 140
 DB 361 GCTAATTATGTTTCACAGCAAGAGCTGAGCTAGGAGTTTATGCAGATGTTGGAAATAA 420
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
 DB 421 ACCTGCGCAGGCTTCCTGGAGCTTTTGGATACCTAGCACATTGATGCCAGACCTTTGCT 480
 QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
 DB 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTGTGTACTGTGACAGTGTGGAAATTTG 540
 QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
 DB 541 GCAGATGGTTATAGCACATGCTCTGGCCCTGATAGGACTGGCAGAGCAATGTGTAC 600
 QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
 DB 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
 QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAGATATAAG 720
 QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 DB 721 AGTATCTTGACCTGGACATCTTTTAAACCAGAGAGAAATTTGTTGATGTTGTGGACCAAGG 780
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 DB 781 GGTGTGAATGACCCAGATATGTTAGTGTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 DB 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCCTAATGACCTC 900
 QY 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
 DB 901 CGACATCAGCCCTCAGCCAAAGCTCTCTTCAGGATAGGACGTAATTCATCATCAAT 960
 QY 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnGlyValTrp 340
 DB 961 CAGGACCCCTTGGGCAAGCAAGGTTACCAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
 QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 DB 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGTATAAACCCGCGAGGAGATTGGT 1080
 QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 DB 1081 GGAACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGTAAAGAGTGGCTGTATTCCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
 DB 1141 GCCTGCTTCATCACACAGCTCTCTCCCTGTGTAAGGAAGCTAGGGTTCCTATGATGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417

```

Db      1201 TCAAGTTAAGAGTCACATAAATCCACAGGCATGTTTGGTTCAGCTA 1251
RESULT 5
ID      AEA27448
XX      AEA27448 standard; DNA; 1254 BP.
XX      AC      AEA27448;
XX      11-AUG-2005 (first entry)
XX      DE      Human alpha-galactosidase DNA, rGAL-12, SEQ ID NO: 15.
XX      KW      Gauchers disease; metabolic; neurological disease; niemann pick disease;
XX      KW      genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
XX      KW      antileptic; cns-gen.; lysosome storage disease; alpha-galactosidase;
XX      KW      gene; ds.
XX      OS      Homo sapiens.
XX      FH      Key      Location/Qualifiers
XX      FT      CDS      1..1254
XX      FT      FT      /*tag= a
XX      FT      FT      /product= "Human alpha-galactosidase protein"
XX      FN      US2005125859-A1.
XX      PD      09-JUN-2005.
XX      XX      08-NOV-2004; 2004US-00984389.
XX      PF      26-JUL-2000; 2000US-00626127.
XX      PR      13-NOV-2001; 2001US-00993059.
XX      PR      20-MAR-2002; 2002US-00103327.
XX      XX      (LARG-) LARGE SCALE BIOLOGY CORP.
XX      PI      Garger SJ, Turpen TH, Kumagai MH;
XX      DR      WPI; 2005-404004/41.
XX      DR      P-PSDB; AEA27449.
XX      PT      New isolated polypeptides useful for producing lysosomal enzymes in
XX      PT      plants to be utilized in enzyme replacement therapy or for the
XX      PT      therapeutic treatment of human or animal lysosomal storage diseases, e.g.
XX      PT      Gaucher's disease.
XX      PS      Disclosure; SEQ ID NO 15; 88pp; English.
XX      CC      The present invention relates to the production of human and animal
XX      CC      lysosomal enzymes in plants by a transient plant expression system. The
XX      CC      invention relates to glucocerebrosidase (GCB, GCR) and alpha-
XX      CC      galactosidase (Gal) enzymes having a post-translational modification
XX      CC      provided by the plant expression system. The invention is useful in
XX      CC      enzyme replacement therapy for treating lysosomal storage diseases such
XX      CC      as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
XX      CC      disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
XX      CC      also useful in researches for developing new approaches to medical
XX      CC      treatment of lysosomal storage diseases and in industrial processes
XX      CC      involving enzymatic substrate hydrolysis. The present sequence is the
XX      CC      human alpha-galactosidase DNA.
XX      SQ      Sequence 1254 BP; 316 A; 289 C; 328 G; 321 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      2,34e-235      Length:      1254
Score:          2293.00      Matches:      417
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     100.00%      Indels:      0
DB:             14      Gaps:        0

US-10-602-220-16 (1-417) x AEA27448 (1-1254)

```

```

1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
1 ATGCAAGCTGAGGAAACCAGAACTACATCTGGGCTGGCGCTTGGCGCTTCGCTTCCTGGCC 60
21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
61 CTCGTTTCTCGGACATCCTCGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCCT 120
41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
121 ACCATGGGCTGGCTGCACCTGGAGCGCTTCATGTGCACTTGACTGCCAGGAAGAGCCA 180
61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
181 GATTCTCGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
241 TGGNAGGATGCAGGTTATGAGTACCTCTGCATTTGATGACTGTTGGATGGCTCCCAAGA 300
101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
301 GATTTCAGAAAGGCAGACTTCAGGCAGACCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
361 GCTAATTATGTTCCACAGCAAGAGCTGAAGCTTAGGATTTATGCAGATGTTGGAAATAA 420
141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
421 ACCTGCGCAGGCTTCCTCGGAGTTTGGATACTACGACATTCATGATGCCAGACCTTTGCT 480
161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
541 GCNAGTGGTTATTAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGNAGCATTTGTGTAC 600
201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATATACAGAAATCCGA 660
221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
661 CAGTACTGCAATCCTGCGGAAATTTTGCTGACATTCATGATGATTCCTGGAAAGTATAAG 720
241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
721 AGTATCTTGACCTGGACATCTTTTAAACCAGGAGAGAAATTTGTGATGTTGCTGGACCAAG 780
261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
781 GGTGTGAATGACCCAGATATGTTAGTATTGGCAACTTTTGGCCTCAGCTGGAATCAGCAA 840
281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTATGATCTAATGACCTC 900
301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCATCAATCAAT 960
321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
961 CAGGACCCCTTGGGCAAGCAAGGCTACCAAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
1021 GAACGACCTCTCTCAGGCTTAGCGCTTGGCTGTAGCTATGATTAACCCGCGAGGATTTGGT 1080

```

QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValIleAlaCysAsnPro 380
DB 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCTCGGTAAAGGAGTGGCCTGTAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValIleArgLysLeuGlyPheTyrGluTyrThr 400
DB 1141 GCTCTGTTCTATCACAGCTCCCTCTGTGTAAGGAGCTAGGGTCTATGATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
DB 1201 TCAAGGTGAAGAAGTCACTAAATCCACAGGCACTGTTTGTCTTCAGCTA 1251

RESULT 6

AAD45223
ID AAD45223 standard; DNA; 1266 BP.
XX AC AAD45223;
XX AC AAD45223;
DT 27-DEC-2002 (first entry)
XX DE Human rGAL-8 DNA.
XX KW Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
XX KW therapeutic; rGAL-8; gene; ds.
XX OS Homo sapiens.

XX Key Location/Qualifiers
XX CDS 1..1266
XX FT /*tag= a
XX FT /product= "Human rGAL-8 protein"

US2002088024-A1.

04-JUL-2002.

PF 13-NOV-2001; 2001US-00993059.

PR 26-JUL-2000; 2000US-00626127.

XX (GARG/) GARGER S J.
XX (TURP/) TURPEN T H.
XX (KUMAI/) KUMAGAI M H.

PI Garger SJ, Turpen TH, Kumagai MH;

DR WPI: 2002-681656/73.

DR P-PSDB; AAE28210.

XX Novel human alpha-galactosidase polypeptide useful for treating lysosomal storage diseases.

PS Claim 1; Page 42-44; 88pp; English.

XX The invention relates to human alpha-galactosidase truncated at the carboxy terminus and the production of enzymatically active recombinant human and animal lysosomal enzymes. The invention is useful for producing CC lysosomal enzymes for treating lysosomal storage diseases, producing CC altered or mutated proteins, enzymatically active or otherwise, to serve CC as precursors or substrates for further in vivo or in vitro processing to CC a specialised industrial form for research or therapeutic uses, to CC produce more effective therapeutic enzyme, for producing antibodies CC against lysosomal enzymes for medical diagnostic use, and in any CC commercial process that involves substrate hydrolysis. The present CC sequence is human rGAL-8 DNA

SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2 37e-235 Length: 1266
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.0% Indels: 0
DB: 6 Gaps: 0
US-10-602-220-16 (1-417) x AAD45223 (1-1266)
QY 1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTCGCTTCCTCTCGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTTCTGGGACATCTCTGGGCTAGAGCACTGGACATGATGTGGCAGAGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGGCTGGCTGCATCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCCA 180
QY 61 AspSerCysIleSerGlnLysLeuPheMetGlnMetAlaGlnLeuMetValSerGluGly 80
DB 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGAAGGATGCAGTTATGAGTACCTCTGCATTTGATGACTGTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTCAGAAAGGCAGACTTCAGGCAGACCTTCAGCGCTTCTCTCATGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 361 GCTAATTATGTTCACACAGCAAGGACTGAAGCTAGGATTTATGCAGATGTGGAAAAATAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCAGGCTTCCTGGGAGTTTGGATACTAGACATGTGATGCCCCAGACTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGAGTAGACTCTGCTAAATTTGATGGTTGTACTGTGACAGATTTGGAAAAATTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 GCAGATGGTTATAGCCACATGTCTTGGCCCTGTAATAGACTGGCAGAGACATTTGTGTAC 600
QY 201 SerCysGluTyrProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
DB 601 TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAGCCCAATTTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGCTGACATTTGATGATTCCTGGAAAAAGTATAAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnIleArgIleValAspValIleAlaGlyProGly 260
DB 721 AGTATCTTGGACTGGACATCTTTTAACCCAGAGAGAAATTTGTGATGTTGCTGCACAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GGTGGATGACCCAGATATGTTAGTTGGCACTTTGGCTTCAGCTCGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACCTCAGATGGCCTCTCGGCTATCATGGCTGCTCTTTTATTCATGTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATTCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
DB 961 CAGGACCCCTTGGGCAAGCAAGGTACCGTCTAGACGGAGACAACTTTTGAAGTGTGG 1020

QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 |||||
 Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCGGCAGAGATTGGT 1080
 |||||
 QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 |||||
 Db 1081 GGACCTCGCTCTTATACCATCCAGTTGCTTCCCTGGGTAAAGGAGTGGCCGTGAATCCT 1140
 |||||
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
 |||||
 Db 1141 GCCTGCTTCATCACAGCTCCTCCTCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
 |||||
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
 |||||
 Db 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTGCTTCAGCTA 1251
 |||||

RESULT 7

ADD84750

ID ADD84750 standard; DNA; 1266 BP.

AC XX

XX ADD84750;

XX 29-JAN-2004 (first entry)

DT XX

DE XX Human alpha-galactosidase rGAL-8 DNA.

XX

KW Human; alpha-galactosidase; rGAL-4; gene; ds; lysosomal enzyme;

KW enzyme replacement therapy; lysosomal disease.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT 1. .1266

FT /*tag= a

FT /product= "Human rGAL-8"

XX

XX US2003106095-A1.

XX

XX 05-JUN-2003.

XX

XX 20-MAR-2002; 2002US-00103327.

XX

XX 26-JUL-2000; 2000US-00626127.

XX

XX 13-NOV-2001; 2001US-00993059.

XX

XX (GARG/) GARGER S J.

XX (TURP/) TURPEN T H.

XX (KUMAI/) KUMAGAI M H.

XX

XX Garger SJ, Turpen TH, Kumagai MH;

XX

XX WPI; 2003-801257/75.

XX

XX P-PSDB; ADD84751.

XX

XX New polynucleotide for producing active recombinant human and animal

XX lysosomal enzymes in a plant expression system that can be used in enzyme

XX replacement therapy.

XX

XX Claim 1; SEQ ID NO 11; 77pp; English.

XX

XX The invention relates to human alpha-galactosidase derivatives and the

XX nucleic acids encoding them. The polypeptides are used in a method for

XX producing active recombinant human and animal lysosomal enzymes in a

XX plant expression system. The enzymes can be used in enzyme replacement

XX therapy for the therapeutic treatment of human and animal lysosomal

XX diseases. This sequence represents DNA encoding a human alpha-

XX galactosidase derivative polypeptide of the invention.

XX

XX SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;

XX

XX Alignment Scores:

XX Pred. No.:

XX Score:

2.37e-235

2293.00

Length:

Matches:

1266

417

Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-602-220-16 (1-417) x ADD84750 (1-1266)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 |||||
 Db 1 ATGCAGCTGAGGAACCCAGAACTATACCTGGGCTGGCGCTTCGCTTCCTCCCTGGCC 60
 |||||
 QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 |||||
 Db 61 CTCGTTTCTCGGACATCCCTGGGGCTAGAGCACTGGACAATGGATGGCAAGGACGCT 120
 |||||
 QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 |||||
 Db 121 ACCATGGGCTGGCTGCACCTGGGAGGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
 |||||
 QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 |||||
 Db 181 GATTCCTGCATCAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAAGC 240
 |||||
 QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 |||||
 Db 241 TGGAGGATGCAGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
 |||||
 QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 |||||
 Db 301 GATTTCAGAGGCGAGACTTCAGGCAGACCTTCAGCGCTTCCTCATGGGATTCGCCAGCTA 360
 |||||
 QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 |||||
 Db 361 GCTAATTATGTTTCACAGCAAGGAGCTGAAGCTAGGAGTTATGACATGTGGCAATATAA 420
 |||||
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
 |||||
 Db 421 ACCTGGCAGGCTTCCTGGGAGTTTGGATACTACACATGTATGCCAGACCTTTGCT 480
 |||||
 QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 |||||
 Db 481 GACTGGGAGTAGACTCTGCTAAAAATTTGATGGTTGTACTGTGACACTTTGGAAAAATTTG 540
 |||||
 QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 |||||
 Db 541 GCAGATGGTTATAAGCACATGCTTCCTGGCCCTGAATAGGACTGGCAGAGCAATGTGTAC 600
 |||||
 QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
 |||||
 Db 601 TCTCTGTAGTGGCTCTTTATATGTGGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
 |||||
 QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 |||||
 Db 661 CAGTACTGCAATCAGCTGGCGCAAAATTTTGTCTGACATTTGATGATTCCTGGAAAAAGTATAAG 720
 |||||
 QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 |||||
 Db 721 AGTATCTTGGACTGGACATCTTTTAAACCAGAGAGAAATTTGTTGATGTTCGTGCACAGGG 780
 |||||
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 |||||
 Db 781 GGTGGAAATGACCCAGATATGTTAGTATGGCACTTTGGCTTCAGCTGGAATCAGCAA 840
 |||||
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 |||||
 Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTCTAATGACCTC 900
 |||||
 QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 |||||
 Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATTAAGACGCTAATGTCATCAAT 960
 |||||
 QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnGluValTrp 340
 |||||

Db 961 CAGGACCCCTTGGCAGCAGGAGGTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
 Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTACCTATGATAAACCGCAGGAGATTGGT 1080
 Qy 361 GlyProArgSerThrIleAlaValAlaSerLeuGlyIysGlyValAlaCysAsnPro 380
 Db 1081 GGACCTCGCTCTTATACCATGCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAATCCT 1140
 Qy 381 AlaCysPheIleThrGlnLeuLeuProValIysArgLysLeuGlyPheTyrGluTrpThr 400
 Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAAGGAGCTAGGGTTCTATGAATGGACT 1200
 Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
 Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTTGCTTCAGCTA 1251
 RESULT 8
 ADJ88276
 ID ADJ88276 standard; DNA; 1266 BP.
 XX AC ADJ88276;
 XX DT 06-MAY-2004 (first entry)
 XX DE Human WT rGAL-8 (galactosidase) DNA.
 XX KW Galactosidase; GAL; gene therapy; lysosomal storage disease;
 XX KW Fabry's disease; Gaucher's disease; human; gene; ds.
 XX OS Homo sapiens.
 XX PH Key Location/Qualifiers
 XX FT CDS 1..1266
 XX FT /*cag= a
 XX FT /product= "Human WT rGAL-8 protein"
 XX PN US2004016021-A1.
 XX PD 22-JAN-2004.
 XX PF 23-JUN-2003; 2003US-00602219.
 XX PR 26-FEB-1988; 88US-00160766.
 XX PR 26-FEB-1988; 88US-00160771.
 XX PR 15-JUL-1988; 88US-00219279.
 XX PR 17-FEB-1989; 89US-00310881.
 XX PR 05-MAY-1989; 89US-00347637.
 XX PR 08-JUN-1989; 89US-00363138.
 XX PR 22-OCT-1990; 90US-00600244.
 XX PR 16-JAN-1991; 91US-00641617.
 XX PR 26-JUL-1991; 91US-00737899.
 XX PR 01-AUG-1991; 91US-00739143.
 XX PR 31-JUL-1992; 92US-00923692.
 XX PR 30-DEC-1992; 92US-00997733.
 XX PR 29-DEC-1993; 93US-00176414.
 XX PR 19-JAN-1994; 94US-00184237.
 XX PR 14-OCT-1994; 94US-00324003.
 XX PR 21-MAY-1999; 99US-00316572.
 XX PR 26-JUL-2000; 2000US-00626127.
 XX PR 13-NOV-2001; 2001US-00993059.
 XX (TURP/) TURPEN T H.
 XX (FOGU/) FOGUE G P.
 XX (ERWI/) ERWIN R L.
 XX (GRIL/) GRILL L K.
 XX Turpen TH, Pogue GP, Erwin RL, Grill LK;
 XX WPI; 2004-108227/11.
 XX P-PSDB; ADJ88277.
 XX

PT New lysosomal enzymes, useful in treating human and animal lysosomal
 PT storage diseases, e.g. Fabry's disease and Gaucher's diseases.
 XX Claim 1; SEQ ID NO 11; 71pp; English.
 XX CC The invention relates to nucleotide encoding galactosidase (GAL). The
 CC invention is useful in gene therapy. The polynucleotides and polypeptides
 CC are useful in treating human and animal lysosomal storage diseases, e.g.
 CC Fabry's disease and Gaucher's diseases. The present sequence is human GAL
 CC DNA.

XX Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,37e-235 Length: 1266
 Score: 2293.00 Matches: 417
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 12 Gaps: 0

US-10-602-220-16 (1-417) x ADJ88276 (1-1266)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTCGGCC 60
 Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 Db 61 CTGCTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACANTGGATTGGCAGGACGCT 120
 Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 Db 121 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
 Qy 61 AspSerCysIleSerGluIysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db 181 GATTCTCTGCATCAGTGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGGC 240
 Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 241 TGGAGAGATGCAGGTATGATGATACCTCTGGATCTGATGATGATGGTGGTCCCAAGA 300
 Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 Db 301 GATTTCAGAGGCGAGACTTCAGGCGAGCCCTCAGCGCTTCTCATGGATTCGCCAGCTA 360
 Qy 121 AlaAsnTyrValHisSerIysGlyLeuIysLeuGlyIleTyrAlaAspValGlyAsnIys 140
 Db 361 GCTAATTTATGTTCCACAGCAAGAGACTGAAGCTAGGATTTATGCAGATGTTGGAATAAA 420
 Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
 Db 421 ACCTGGCAGGCTTCCCTGGAGTTTGGATATACGACATTCATGCCAGAGCTTTGCT 480
 Qy 161 AspTrpGlyValAspLeuLeuIysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAAATTG 540
 Qy 181 AlaAspGlyTyrIysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 Db 541 GCAGATGGTTATTAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
 Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnIysProAsnTyrThrGluIleArg 220
 Db 601 TCTGTGAGTGGCTCTTTATATGTGGCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
 Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpIysSerIleLys 240
 Db 661 CAGTACTGCAATCACTGGCGAATTTTGTGACATTTGATGATGATCTCTCTGGAAGATATAAG 720
 Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260

```
Db 721 AGTATCTTGGCTGGAGCATCTTTTAAACACGAGAGAGAAATGTTGATGTTGCTGACCAAGG 780
Qy 261 GlyTyrAsnAspProAspMetIleuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAATGACCAATATGTTAGTGATGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCGGGCTATCATGGCTGCTCCTTTATTCATGCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaIysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGCTAATTTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTTCAGCTTAGCTGGCTGTAGCTATGATAAACCGCAGGAGATGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTGGGTAAGGAGTGGCCTGTAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTGGCTTCAGCTA 1251
RESULT 9
ID ADM48680
XX ADM48680 standard; DNA; 1266 BP.
AC ADM48680;
XX
XX 03-JUN-2004 (first entry)
XX Human wild type rGAL-8 DNA.
XX Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
KW gene; ds.
XX
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1266
FT FT /*tag= a
FT FT /product= "GAL-8 protein"
XX
FN US2004023281-A1.
XX
XX 05-FEB-2004.
XX
XX 23-JUN-2003; 2003US-00602220.
XX
XX 26-FEB-1988; 88US-00160766.
XX 26-FEB-1988; 88US-00160771.
XX 17-FEB-1989; 89US-00310881.
XX 22-OCT-1990; 90US-00600244.
XX 31-JUL-1992; 92US-00923692.
XX 30-DEC-1992; 92US-00997733.
XX 28-DEC-1993; 93US-00176414.
XX 19-JAN-1994; 94US-00184237.
XX 14-OCT-1994; 94US-00324003.
XX 21-MAY-1999; 99US-00316572.
XX 26-JUL-2000; 2000US-00626127.
```

```
PR 13-NOV-2001; 2001US-00993059.
XX
XX (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
PA (FOGU/) FOGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX
XX Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
XX
XX WPI; 2004-142650/14.
XX P-PSDB; ADM48681.
XX
XX New alpha-galactosidase polypeptides, useful in producing recombinant
XX lysosomal enzymes for the treatment of lysosomal storage diseases, such
XX as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
XX disease.
XX
XX Disclosure; SEQ ID NO 11; 72pp; English.
XX
XX The present invention relates to novel galactosidase (Gal) proteins such
XX as rGAL-12, rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions
XX of the present invention are useful for producing recombinant lysosomal
XX enzymes for enzyme replacement therapy for treating human and animal
XX lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
XX disease, Fabry disease and Tay-Sachs disease. The present sequence is
XX human wild type rGAL-8 DNA used in the exemplification of the invention.
XX
XX Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 2,37e-235 Length: 1266
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0
US-10-602-220-16 (1-417) x ADM48680 (1-1266)
Qy 1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCTTGCCTTGCCTTCTTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGTTTCTTGGACATCTCTGGGGCTAGAGCACTGGACATGATGGTGGCAAGAGCGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Qy 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTTCATCATCAGTGAAGAGCTTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysAlaAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGAGATGCAGGCTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAAGAGGCAGACTTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATGTTTACACAGCAAGAGACTGAAAGCTAGGGAATTTATGCATGTGTGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTTGGATATACGACATTCATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
```

```
Db 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGACAGTTTGGAAAATTTG 540
Qy 181 AlaAspGlyTyrIysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAACATTTGTATC 600
Qy 201 SerCysGluTrpProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCCCTCTTTATATATGTGGCCCTTTCAAAAGGCCAATTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspSerTyrIysSerIleLys 240
Db 661 CAGTACTGCAATCACTGCGGAAATTTTGTGACATGATTCCTCGAAAGATATAAAG 720
Qy 241 SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGAGTGCATCTTTTAAACGAGGAGAAATTTGTGATGTGCTGGACACGG 780
Qy 261 GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGACATATGTTAGTGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTyrAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTCTCTTTATTTCTATGATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATTAAGGACGTAATTTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnGluValTyr 340
Db 961 CAGGACCCCTTGGCAAGCAAGGTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCGCTGTAGTGTATGATTAACCGCGCAGAGATGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTCTCTGGTAAAGAGTGGCTGTATCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db 1141 GCCTGCTTTCATCACACAGCTCTCTCTGTGAAAGGAAGCTAGGGTTCTATGAATGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAAGTCACATAAATCCACAGGCACTGTTTTGTCTTCAGCTA 1251

RESULT 10
ADU66915
ID ADU66915 standard; DNA; 1266 BP.
AC
XX ADU66915;
XX AC
XX 10-FEB-2005 (first entry)
XX
XX Human alpha-galactosidase protein encoding DNA #5.
XX
XX Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase;
KW enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;
KW Niemann-Pick disease; Fabry's disease; Tay-Sachs disease;
KW Hurler's syndrome; Hurler-Scheie syndrome; nephrotropic; human; gene; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX 1. .1266
XX /*tag= a
XX /product= "Human alpha-galactosidase protein"
```

```
PN US2004234516-A1.
XX
PD 25-NOV-2004.
XX
XX 21-MAY-2004; 2004US-00851388.
XX
XX 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
PR 20-MAR-2002; 2002US-00103327.
XX
XX (LARG-) LARGE SCALE BIOLOGY CORP.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
XX
XX WPI; 2004-821274/81.
XX
XX P-PSDB; ADU66916.
XX
XX A pharmaceutical composition comprising a lysosomal enzyme, useful for
XX enzyme replacement therapy for the treatment of lysosomal storage
XX diseases, such as Fabry's disease.
XX
XX Disclosure; SEQ ID NO 11; 86pp; English.
XX
XX The present invention relates to the production of human and animal
XX lysosomal enzymes in plants by a transient plant expression system. The
XX invention relates to glucocerebrosidase (GCB, GCR) and alpha-
XX galactosidase enzymes having a post-translational modification provided
XX by the plant expression system. The invention is useful in enzyme
XX replacement therapy for treating lysosomal storage diseases such as
XX Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs
XX disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
XX also useful in researches for developing new approaches to medical
XX treatment of lysosomal storage diseases and in industrial processes
XX involving enzymatic substrate hydrolysis. The present sequence is the
XX human alpha-galactosidase protein encoding DNA.
XX
XX SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,37e-235 Length: 1266
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 13 Gaps: 0

US-10-602-220-16 (1-417) x ADU66915 (1-1266)

Qy 1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTCTGCGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGCTTCTCTGGACATCCCTGGGGCTAGAGACTGGACATGGACATGGATTTGGCAGACGCT 120
Qy 41 ThrMetGlyTyrTrpLeuHisTyrGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCTGCTGAGGAGCGCTTCATGTGCAACCTTCACTGCCAGGAAGACCA 180
Qy 61 AppSerCysIleSerGlnLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTTCTGTCATCATGTAGAGAGCTTTTCATGGAGATGCGAGAGCTCATGGTCTCAGAAAGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTyrTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGCGAGTTATGATGATACCTCTGCAATTGATGACTGTTGGTGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAGGACAGACTTCAGGACAGACCTTCAGCGCTTTCTCATGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
```

Db 361 GCTAATTATGTTACAGCAAGGACTGAGCTAGGATTTATGACATGTTGGAATATA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCTCGGAGTTTGGATACATGATGATGATGATGATGATGATGATGAT 480
Qy 161 AspTyrGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGACATGCTCTTGGCCCTGATAGGATGCTGCGAAGCATTTGTGATC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGCTGCATATTGATGATTCCTGGAAAGTATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGACTGGACATCTTTTAACAGGAGAGATTTGTTGATGTTGCTGGACCCAGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTATTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCATAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCATGACCCCTCAAGCAAGGCTCTCTCTCAGGATNAGGACGTAATTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGCAGCAAGGCTACCAAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTTAGCTGGCTGTAGCTATGATGATAAACCGGACGAGATGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCTCTGGTAAAGAGTGGCTGTATCTCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCTCTCTCTGTGAAAGGAAGCTAGGGTTCTATGAATGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGTGTTCTAGCTA 1251
RESULT 11
AEA27444
ID AEA27444 standard; DNA; 1266 BP.
XX AC AEA27444;
XX AC AEA27444;
XX 11-AUG-2005 (first entry)
XX Human alpha-galactosidase DNA, rGAL-8, SEQ ID NO: 11.
XX Gauchers disease; metabolic; neurological disease; niemann pick disease;
KW genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
KW antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase;
KW gene; ds.

XX Homo sapiens.
OS XX
XX Key Location/Qualifiers
CDS 1..1266
/*tag= a
/product= "Human alpha-galactosidase protein"
XX US2005125859-A1.
XX 09-JUN-2005.
XX 08-NOV-2004; 2004US-00984389.
XX 26-JUL-2000; 2000US-00636127.
XX 13-NOV-2001; 2001US-00993059.
XX 20-MAR-2002; 2002US-00103327.
XX (LARG-) LARGE SCALE BIOLOGY CORP.
XX Garger SJ, Turpen TH, Kumagai MH;
PI P-PSDB; AEA27445.
XX WPI; 2005-404004/41.
XX DR P-PSDB; AEA27445.
XX PT New isolated polypeptides useful for producing lysosomal enzymes in
PT plants to be utilized in enzyme replacement therapy or for the
PT therapeutic treatment of human or animal lysosomal storage diseases, e.g.
PT Gaucher's disease.
XX Disclosure; SEQ ID NO 11; 88pp; English.
XX The present invention relates to the production of human and animal
XX lysosomal enzymes in plants by a transient plant expression system. The
XX invention relates to glucocerebrosidase (GCB, GCR) and alpha
XX galactosidase (Gal) enzymes having a post-translational modification
XX provided by the plant expression system. The invention is useful in
XX enzyme replacement therapy for treating lysosomal storage diseases such
XX as Gaucher's disease, Niemann-pick disease, Fabry's disease, Tay-Sachs
XX disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
XX also useful in researches for developing new approaches to medical
XX treatment of lysosomal storage diseases and in industrial processes
XX involving enzymatic substrate hydrolysis. The present sequence is the
XX human alpha-galactosidase DNA.
XX Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 2,37e-235 Length: 1266
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0
US-10-602-220-16 (1-417) x AEA27444 (1-1266)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTCGCTTCCTCTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyValaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCGGACATCCTCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTTCATCATGTAGAAGCTCTTCATGGAGATGGCAGAGCTCATGTGTTCTAGAAGGC 240


```
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCTCTGCATCAGTCAGAGAGCTCTTCATGGAGATGGCAGAGCTCAATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGGNAGGATGCAGTTATGATACCTCTGCATTTGATGACTGTTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTCAAGAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 361 GCTAAATATGTTTCCACAGCAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAAATAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACTACGACATTGATGCCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerIleGluAsnLeu 180
DB 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTACACATTTTGGAAAATTTG 540
QY 181 AlaAspGlyTyrIlyshMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 GCAGATGGTTATTAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCATGTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
DB 601 TCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAGGCCAATTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTGATGATTCCTGGAAAAGTATAAAG 720
QY 241 SerIleAspTrpThrSerPheAsnGlnGluArgIleValAspAlaGlyProGly 260
DB 721 AGTATCTTGACATGGACATCTTTAAACCAGGAGAGAAATTTGATGTGCTGGCAGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GGTGGNATGACCCAGATATGTTAGTGTGATGGCACTTTGGCTCAGCTGGAATCAGCA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATTAAGGACGTAATGGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTTGGCAGCAGGAGTACAGCTTAGACAGGAGCAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATAAACCCGCGAGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerIleuGlyLysGlyValAlaCysAsnPro 380
DB 1081 GGAACCTCGCTTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValIleArgLysLeuGlyPheTyrGluTrpThr 400
DB 1141 GCCTGCTTCATCACAGACTCTCTCCCTGTGAAAAGGAGCTAGGGTTCTATGATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
DB 1201 TCAAGGTTAAGAGTACATAAAATCCACAGGCACCTGTTTTGCTTCAGCTA 1251
```

RESULT 13

```
ADD84756
ID ADD84756 standard; DNA; 1272 BP.
XX AC ADD84756;
XX 29-JAN-2004 (first entry)
XX DE Human alpha-galactosidase rGAL-12R DNA.
XX KW Human; alpha-galactosidase; rGAL-4; gene; ds; lysosomal enzyme;
XX KW enzyme replacement therapy; lysosomal disease.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT CDS l..1272
FT /tag= a
FT /product= "Human rGAL-12R"
XX PN US2003106095-A1.
XX PD 05-JUN-2003.
XX PF 20-MAR-2002; 2002US-00103327.
XX PR 26-JUL-2000; 2000US-00626127.
XX PR 13-NOV-2001; 2001US-00993059.
XX PA (GARG/) GARGER S J.
XX PA (TURP/) TURPEN T H.
XX PA (KUMA/) KUMAGAI M H.
XX PI Garger SJ, Turpen TH, Kumagai MH;
XX DR WPI; 2003-801257/75.
XX DR P-PSDB; ADD84757.
XX PT New polynucleotide for producing active recombinant human and animal
XX PT lysosomal enzymes in a plant expression system that can be used in enzyme
XX PT replacement therapy.
XX PS Claim 1; SEQ ID NO 17; 77pp; English.
XX CC The invention relates to human alpha-galactosidase derivatives and the
XX CC nucleic acids encoding them. The polypeptides are used in a method for
XX CC producing active recombinant human and animal lysosomal enzymes in a
XX CC plant expression system. The enzymes can be used in enzyme replacement
XX CC therapy for the therapeutic treatment of human and animal lysosomal
XX CC diseases. This sequence represents DNA encoding a human alpha-
XX CC galactosidase derivative polypeptide of the invention.
XX SQ Sequence 1272 BP; 323 A; 291 C; 333 G; 325 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,39e-235 Length: 1272
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-602-220-16 (1-417) x ADD84756 (1-1272)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTGTTTCTCTGGGACATCCCTCGGGCTAGAGCACTGGACATGGATGGCAGGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
```

Db 121 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTCCAGGAGAGCCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCTGCATCAGTGAAGAGCTCTTCATGAGATGGCAGAGCTCATGTGCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGGATCAGGTTATGATGATCCTCTGTCATTTGATGACTGTTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGCAGACCTTCAGCAGACCTTCAGGCGCTTCTCATGAGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGGACTGAAGCTAGGATTTATGTCAGATGTTGGAATATAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACCTACGACATTGATGCCAGACTTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTCTGACAGTTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATTAAGCACATGCTCTTGCCCTCGAATAGGACTGGCAGAACCAATTGTGAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProLeuTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGGCCAATTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCAATTGATTCCTGGAAAGATATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGACTGGACATCTTTTAACCAAGAGAGAAATGCTGATGTTGTCGACCAAGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTTCGAATCACCAGATATGTTAGTGTGGAACCTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCAGCCCTCAAGCAAAAGCTCTCTTCAGGATAAGGACGTAATTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGCAAGCAAGGCTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGGCTGGCTGTGATGATGATGATGATGATGATGATGAT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGAGTGGCTGTATCTCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTGAAAGGAAGCTAGGCTTCTATGATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTTGCTTCAGCTA 1251

RESULT 14

ADJ88282

ID ADJ88282 standard; DNA; 1272 BP.

XX

AC ADJ88282;

XX

DT 06-MAY-2004 (first entry)

XX

DE Human WT rGAL-12R (galactosidase) DNA.

XX

KW Galactosidase; GAL; gene therapy; lysosomal storage disease;

KW Fabry's disease; Gaucher's disease; human; gene; ds.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT CDS 1..1272

FT

FT /tag= a

FT /product= "Human WT rGAL-12R protein"

XX

PN US2004016021-A1.

XX

PD 22-JAN-2004.

XX

PF 23-JUN-2003; 2003US-00602219.

XX

PR 26-FEB-1988; 88US-00160766.

PR

PR 26-FEB-1988; 88US-00160771.

PR

PR 15-JUL-1988; 88US-00219279.

PR

PR 17-FEB-1989; 89US-00310881.

PR

PR 05-MAY-1989; 89US-00347637.

PR

PR 08-JUN-1989; 89US-00363138.

PR

PR 22-OCT-1990; 90US-00600244.

PR

PR 16-JAN-1991; 91US-00641617.

PR

PR 26-JUL-1991; 91US-00737899.

PR

PR 31-JUL-1992; 92US-00923692.

PR

PR 30-DEC-1992; 92US-00997733.

PR

PR 29-DEC-1993; 93US-00176414.

PR

PR 19-JAN-1994; 94US-00184237.

PR

PR 14-OCT-1994; 94US-00324003.

PR

PR 21-MAY-1999; 99US-00316572.

PR

PR 26-JUL-2000; 2000US-00626127.

PR

PR 13-NOV-2001; 2001US-00993059.

XX

(TURP/) TURPEN T H.

PA

(POGU/) POGUE G P.

PA

(ERWL/) ERWIN R L.

PA

(GRIL/) GRILL L K.

XX

Turpen TH, Pogue GP, Erwin RL, Grill LK;

PI

XX

DR

DR

XX

XX

XX

XX

XX

XX

Alignment Scores:

Pred. No.: 2,39e-235

Score: 2293.00

Percent Similarity: 100.00%

Length: 1272

Matches: 417

Conservative: 0

New lysosomal enzymes, useful in treating human and animal lysosomal storage diseases, e.g. Fabry's disease and Gaucher's diseases.

Claim 1; SEQ ID NO 17; 71pp; English.

The invention relates to nucleotide encoding galactosidase (GAL). The invention is useful in gene therapy. The polynucleotides and polypeptides are useful in treating human and animal lysosomal storage diseases, e.g. Fabry's disease and Gaucher's diseases. The present sequence is human GAL DNA.

Sequence 1272 BP; 323 A; 291 C; 333 G; 325 T; 0 U; 0 Other;

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0

US-10-602-220-16 (1-417) x ADJ88282 (1-1272)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCTTCGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTCTCGGACATCCTCGGGCTAGGCACTGGACAATGATTTGGCAAGGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAAGCTTGACTGCCAGGAAGCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCCTGCATCAGTGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGGAGGATGCAGTTATGATACCTCTGCAATGATGACTGTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTCAGAGGACAGCTTCAGGCAGACCCCTCAGCGCTTCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 361 GCTAATTTATGTTCCACAGCAAGGACTGAAGCTAGGATTTATGCAGATGTTGGAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCAGGCTTCCTCGGAGCTTTGGNATCTACGACATTCATGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGAGTAGATCTGCTAAAAATTTGATGGTTGTACTGTACACATTTGGAAAAATTTG 540
QY 181 AlaAspGlyTyrIlyHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 GCAGATGGTTATGAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
DB 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGGCCAATTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGTGACATTTGATGATTCCTGGAAAAAGTATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 721 AGTATCTGAGCTGGACATCTTTTAAACCAGGAGAAATTTGTATGTTGTGCACACGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GGTGTGNATGACCAGATATGTTAGTGTGATTTGGCACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACTCAGATGGCCCTCTGGGTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGATTAAGGACGTAATTCCTCATCAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTTGGCAAGCAAGGCTACACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020

QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCTCAGGCTTACCTGGGCTGAGCTATGATATAACCGGAGGAGATTGCT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1081 GGACCTCGCTCTATATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
DB 1141 GCCTGCTTTCATCACAGCTCCTCCTGTGAAAGGAAGCTAGGGTCTTATGAATGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
DB 1201 TCAAGGTTAAGAAGTACATAAATCCACAGGCACTGTTTTGCTTCAGCTA 1251

RESULT 15
ADM48686
ID ADM48686 standard; DNA; 1272 BP.
XX
AC ADM48686;
XX
DT 03-JUN-2004 (first entry)
XX
DE Human wild type rGAL-12R DNA.
XX
KW Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
KW gene; ds.
XX
OS Homo sapiens.
XX

Key Location/Qualifiers
FH 1. .1272
FT CDS
FT /*tag= a
FT /product= "GAL-12R protein"
XX
XX US2004023281-A1.
XX
XX 05-FEB-2004.
XX
XX 23-JUN-2003; 2003US-00602220.
XX
XX 26-FEB-1988; 88US-00160766.
XX 26-FEB-1988; 88US-00160771.
XX 17-FEB-1989; 89US-00310881.
XX 22-OCT-1990; 90US-00600244.
XX 31-JUL-1992; 92US-00923692.
XX 30-DEC-1992; 92US-00997733.
XX 29-DEC-1993; 93US-00176414.
XX 19-JAN-1994; 94US-00184237.
XX 14-OCT-1994; 94US-00324003.
XX 21-MAY-1999; 99US-00316572.
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.
XX
XX (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX (POGU/) POGUE G P.
XX (ERWI/) ERWIN R L.
XX (GRIL/) GRILL L K.
XX
XX Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
XX WPI: 2004-142650/14.
XX P-PSDB; ADM48687.
XX
XX New alpha-galactosidase polypeptides, useful in producing recombinant
XX lysosomal enzymes for the treatment of lysosomal storage diseases, such
XX as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
XX disease.

XX Disclosure; SEQ ID NO 17; 72pp; English.
 XX The present invention relates to novel galactosidase (Gal) proteins such
 CC as rGAL-12, rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions
 CC of the present invention are useful for producing recombinant lysosomal
 CC enzymes for enzyme replacement therapy for treating human and animal
 CC lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
 CC disease, Fabry disease and Tay-Sachs disease. The present sequence is
 CC human wild type rGAL-12R DNA. This sequence comprises a human rGAL-12 DNA
 CC and ER retention signal DNA. This sequence is used in the exemplification
 CC of the invention.

XX SQ Sequence 1272 BP; 323 A; 291 C; 333 G; 325 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,39e-235 Length: 1272
 Score: 2293.00 Matches: 417
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 12 Gaps: 0

US-10-602-220-16 (1-417) x ADM48686 (1-1272)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
 QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 DB 61 CTCGTTCCTGGGACATCCCTGGGGCTAGAGCACTGGACATGGATTGGCAAGACGCT 120
 QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 DB 121 ACCATGGGCTGGCTGGAGCGCTTCATGTGCACCTTGACTGCCAGGAGAGCCCA 180
 QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 DB 181 GATTTCCTGCATCAGTGAGAGCTCTTCATGCAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
 QY 81 TrpIysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 DB 241 TGGAAAGATGAGGTATAGTACCTCTGATGATGATGATGATGATGATGATGATGATGAT 300
 QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 DB 301 GATTCAAGGCGAGCTTCAGGCGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
 QY 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
 DB 361 GCTAATTTATGTTACAGCAAAAGGACTGAAGCTAGGATTTATGCAGATGTTGGAAATATA 420
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
 DB 421 ACCTGCGCAGGCTTCCTGGGAGTTTGGATACCTACGACATTCATGCCAGACCTTTGCT 480
 QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
 DB 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 540
 QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
 DB 541 GCAGATGTTATAGCAGCATGCTTCCTGGCCCTGAAATAGGACTGGCAGAGCATTTGTGTAC 600
 QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
 DB 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGGCCCAATTATACAGAAATCCGA 660
 QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 DB 661 CAGTACGCAATCACTGGCGAAATTTTGTGCTGACATTGATGATTCCTGGAAAGTATTAAG 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 DB 721 AGTATCTTGAGCTGGACATCTTTAAACCAGGAGAGAAATTTGATGTGTGACCAAGGG 780
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 DB 781 GGTGGANYGACCAGATATGTAGTGTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
 QY 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
 DB 901 CGACATCATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCGCATCAAT 960
 QY 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 DB 961 CAGGACCCCTTGGGCAAGCANGGGTACCAGCTTAGACAGGGGAGACAACTTTGAAGTGTGG 1020
 QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGCAGGAGATTGGT 1080
 QY 361 GlyProArgSerTrpThrIleAlaValAlaSerIleuGlyLysGlyValAlaCysAsnPro 380
 DB 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTGTAAATCCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
 DB 1141 GCCTGTCTTCATCACAGCTCTCCCTGTGAAAAGGAGGCTAGGGTCTATGATGAGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
 DB 1201 TCAAGGTAAAGAGTCACATAAATCCACAGGCACTGTTTGTGCTTCAGCTA 1251

Search completed: January 1, 2006, 00:49:52
 Job time : 1146.75 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2006, 00:27:04 ; Search time 9516.23 Seconds
(without alignments)
2490.874 Million cell updates/sec

Title: US-10-602-220-16

Perfect score: 2293

Sequence: 1 MQLRNLHLCALRFLA.....EWSRLSRSHNPVTGLLQL 417

Scoring table: BLOSUM62

Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US10602220/runat_30122005_140557_5421/app_query.fasta_1.782
-DBS-GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOFC=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10602220@cgn 1.1 5556 @runat_30122005_140557_5421 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.in.*
- 3: gb.env.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pr.*
- 9: gb.ro.*
- 10: gb.sts.*
- 11: gb.sy.*
- 12: gb.un.*
- 13: gb.vi.*
- 14: gb.htg.*
- 15: gb.pl.*

ALIGNMENTS

RESULT 1	AR653698	AR653698	Sequence 15 from patent US 6887696.	1254 bp	DNA	linear	PAT 13-JUN-2005
LOCUS	AR653698	Sequence 15 from patent US 6887696.					
DEFINITION	AR653698	Sequence 15 from patent US 6887696.					
ACCESSION	AR653698	Sequence 15 from patent US 6887696.					
VERSION	AR653698.1	GI:67584283					
KEYWORDS	Unknown.						
SOURCE	Unknown.						
ORGANISM	Unknown.						
REFERENCE	1 (bases 1 to 1254)						
AUTHORS	Gargner, S.J., Turpen, T.H. and Kumagai, M.H.						
TITLE	Production of lysosomal enzymes in plants by transient expression						
JOURNAL	Patent: US 6887696-A 15 03-MAY-2005;						
FEATURES	Large Scale Biology Corporation; Vacaville, CA						
source	Location/Qualifiers						
source	1..1254						
source	/organism="unknown"						
source	/mol_type="genomic DNA"						
ORIGIN							
Alignment Scores:							
Pred. No.:	1.29e-211	Length:	1254				
Score:	2293.00	Matches:	417				
Percent Similarity:	100.00%	Conservative:	0				
Best Local Similarity:	100.00%	Mismatches:	0				
Query Match:	100.00%	Indels:	0				

DB:	6	Gaps:	0
US-10-602-220-16 (1-417) x AR653698 (1-1254)			
QY	1	MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20
DB	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTCGGCC	60
QY	21	LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro	40
DB	61	CTCGTTTCCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGCGCCT	120
QY	41	ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluPro	60
DB	121	ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCCA	180
QY	61	AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly	80
DB	181	GATTCTCGCATCAGTGAGAGCTTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAGGC	240
QY	81	TrpIleAspAlaGlyTrpGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg	100
DB	241	TGAAGGATGCAGGTATAGTACCTCTGCAATTGATGACTGTGGTGGCTCCCAAGA	300
QY	101	ArgHisIleSerProGlnAlaLeuAlaLeuGlnAspLysAspValIleAlaIleAsn	320
DB	901	CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGCGTAATTGGCCATCAAT	960
QY	321	GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp	340
DB	961	CAGGACCCCTTGGGCAAGCGGTACCGCTTAGACGGGAGACAACCTTTGAAGTGTGG	1020
QY	341	GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly	360
DB	1021	GAAGCACTCTCTCAGGCTTAGCTGGCTAGCTATGATAAACCCGCGAGAGATTGGT	1080
QY	361	GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyGlyValAlaCysAsnPro	380
DB	1081	GGACCTCGCTCTTATACCATCGCAGTGTCTCCCTGGGTAAAGAGTGGCTGTAAATCCT	1140
QY	381	AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr	400
DB	1141	GCCTGCTTTCATCACAGCTCTCCCTGTGAAAGGAAGCTAGGTTCTTATGAATGGACT	1200
QY	401	SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu	417
DB	1201	TCAAGGTAAAGAGTCACATAATCCACAGGCACTGTTTGTCTTCAGCTA	1251
RESULT 2			
AR654966 1254 bp DNA linear PAT 13-JUN-2005			
Sequence 15 from patent US 6890748.			
ACCESSION AR654966			
VERSION AR654966.1 GI:67586505			
KEYWORDS Unknown.			
SOURCE Unknown.			
ORGANISM Unclassified.			
REFERENCE 1 (bases 1 to 1254)			
AUTHORS Garger, S.J., Turpen, T.H. and Kumagai, M.H.			
TITLE Production of lysosomal enzymes in plants by transient expression			
JOURNAL Patent: US 6890748-A 15 10-MAY-2005;			
Large Scale Biology Corporation; Vacaville, CA			
FEATURES			
Location/Qualifiers			
1..1254			
/organism="unknown"			
/mol_type="genomic DNA"			
ORIGIN			
Alignment Scores:			
Pred. No.:	1,296-211	Length:	1254
Score:	2293.00	Matches:	417
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0
US-10-602-220-16 (1-417) x AR654966 (1-1254)			
QY	1	MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20
DB	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTCGGCC	60
QY	21	LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro	40
DB	61	CTCGTTTCCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGCGCCT	120
QY	41	ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluPro	60
DB	121	ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCCA	180
QY	61	AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly	80
DB	181	GATTCTCGCATCAGTGAGAGCTTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAGGC	240
QY	81	TrpIleAspAlaGlyTrpGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg	100
DB	241	TGAAGGATGCAGGTATAGTACCTCTGCAATTGATGACTGTGGTGGCTCCCAAGA	300
QY	101	ArgSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu	120
DB	301	GATTCAAGGCGAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA	360
QY	121	AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys	140
DB	361	GCTAATTATGTTACAGCAAGGACTGAAGCTAGGAGTTATGCAGATGTTGGAATAAA	420
QY	141	ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla	160
DB	421	ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTCATGCCAGACCTTTGCT	480
QY	161	AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu	180
DB	481	GACTGGGAGTAGACTGCTTAAATTTGATGGTTTACTGTGACATTTGGAAATTTG	540
QY	181	AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr	200
DB	541	GCAGATGGTTAAGACACATGCTCTTGGCCCTGAATAGGACTGGCAAGCACTTGTGTAC	600
QY	201	SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg	220
DB	601	TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGGCCCAATTATACAGAAATCCGA	660
QY	221	GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys	240
DB	661	CAGTACTGCAATCACTGGGAAATTTTGTGACATTCATGATTCCTGGAAAGTATTAAG	720
QY	241	SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly	260
DB	721	AGTATCTTGAGCTGGAATCTTTTAAACAGGAGAGAAATGTTGATGTGTGGACAGGG	780
QY	261	GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln	280
DB	781	GGTTGGAATGACCCAGATATGTTAGTATGGCAACTTTGGCCTCAGCTGGAATCAGCAA	840
QY	281	ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu	300
DB	841	GTAATCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCATATGACCTC	900
QY	301	ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn	320
DB	901	CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGCGTAATTGGCCATCAAT	960
QY	321	GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp	340
DB	961	CAGGACCCCTTGGGCAAGCGGTACCGCTTAGACGGGAGACAACCTTTGAAGTGTGG	1020
QY	341	GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly	360


```
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATACATGACATGATGCCAGACCTTTGCT 480

QY 161 AspTyrGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTTACTGTGACAGTTTGGAAATTTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 GCAGATGGTTATAGACACATGCTCTTGGCCCTGATAGGACTGGCAGAACATTTGTAC 600

QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProLeuTyrThrGluLeuArg 220
DB 601 TCCTGTGAGTGGCTCTTTATATATGTCGCCCTTTCAAAAGGCCAATTTATACAGAAATCCGA 660

QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACTGGCGCAATTTTGTGACATTTGATGATTCCTGGAAAGATATAAG 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 721 AGTATCTTGACCTGGACATCTTTTACCAGAGAGATTTGTTGATTTGCTGGACCAAGG 780

QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GGTGGAAATCACCCAGATATGTTAGTTGGCACTTTGGCCCTCAGCTGGAAATCAGCAA 840

QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTATGTTCTAATGACCTC 900

QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACACATCAGCCCTCAGGCCAAAGCTCTCTTCAGGATTAAGACGTTAATGGCATCAAT 960

QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTTGGCAAGCAAGGTCACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020

QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCTCAGCTTTAGCTGGCTGTAGCTATGATATNAACCGGACAGAGATGGT 1080

QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1081 GGACCTCGCTTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTATCCT 1140

QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
DB 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTCAAAAGGAAGCTAGGGTTCTATGAATGACT 1200

QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
DB 1201 TCAAGGTTAAGAGTACATAAAATCCACAGGCACTGTTTGTCTCAGCTA 1251

RESULT 3
AR653696
LOCUS AR653696 1266 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 11 from patent US 6887696.
ACCESSION AR653696
VERSION AR653696.1 GI:67584281
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1266)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 11 03-MAY-2005;
FEATURES Large Scale Biology Corporation; Vacaville, CA
Location/Qualifiers
```

source

1. 1266
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:

Pred. No.: 1.3e-211 Length: 1266
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-16 (1-417) x AR653696 (1-1266)

```
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATCCAGCTGAGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTCGTCCG 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTTCTGGGACATCCCTGGGCTTAGAGCACTGGACAATGGATTGGCAAGACGCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCA 180

QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCCTGCATCAGTGAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240

QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TCGAAGGATGACGGTTATGAGTACCTCTGCAATTCATGACTGTTGGATGGCTCCCAAGA 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTCAGAGGAGAGACTTCAGGAGACCCCTCAGCGCTTTCCTCATGGATTCGCAGCTA 360

QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 361 GCTAATTTATGTTACAGCAAAAGGACTGAAGCTAGGAGATTTATGCAGATTTGGAATAA 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATACATGACATTCATGATGCCAGACCTTTGCT 480

QY 161 AspTyrGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTTACTGTGACAGTTTGGAAATTTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 GCAGATGGTTATAGACACATGCTCTTGGCCCTGATAGGACTGGCAGAACATTTGTAC 600

QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProLeuTyrThrGluLeuArg 220
DB 601 TCCTGTGAGTGGCTCTTTATATATGTCGCCCTTTCAAAAGGCCAATTTATACAGAAATCCGA 660

QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACTGGCGCAATTTTGTGACATTTGATGATTCCTGGAAAGATATAAG 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 721 AGTATCTTGACCTGGACATCTTTTACCAGAGAGATTTGTTGATTTGCTGGACCAAGG 780

QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GGTGGAAATCACCCAGATATGTTAGTTGGCACTTTGGCCCTCAGCTGGAAATCAGCAA 840

QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
```

```

Db      841  GTAACATCAGATGCGCCCTCTGGGCTATCATGGCTGCTCCTTTATTATCATGCTCTAATGACCTC 900
QY      |||
QY      301  ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db      |||
Db      901  CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTAGGATAAGGACGTAATTCGCAAT 960
QY      |||
QY      321  GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db      |||
Db      961  CAGGACCCCTTGGGCAAGCAAGGTACCGCTTAGACAGGAGCAACTTTGAAGTGG 1020
QY      |||
QY      341  GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db      |||
Db      1021  GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCCGCGAGAGATTGGT 1080
QY      |||
QY      361  GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      |||
Db      1081  GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGAGTGGCTGTAATCCT 1140
QY      |||
QY      381  AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db      |||
Db      1141  GCCTGCTTCATCACACAGCTTCCCTGTGAAAGGAAGCTAGGGTTCTATGATGGACT 1200
QY      |||
QY      401  SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db      |||
Db      1201  TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTA 1251

RESULT 4
AR654964
LOCUS      1266 bp      DNA      linear      PAT 13-JUN-2005
DEFINITION      Sequence 11 from patent US 6890748.
ACCESSION      AR654964
VERSION      AR654964.1  GI:67586503
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unknown.
REFERENCE
  1 (bases 1 to 1266)
  Garger, S.J., Turpen, T.H. and Kumagai, M.H.
  Production of lysosomal enzymes in plants by transient expression
  Patent: US 6890748-A 11 10-MAY-2005;
  Large Scale Biology Corporation; Vacaville, CA
FEATURES
  source
    1..1266
    /organism="unknown"
    /mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.:      1.3e-211      Length:      1266
Score:          2233.00      Matches:      417
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:             6           Gaps:      0

US-10-602-220-16 (1-417) x AR654964 (1-1266)
QY      1  MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db      |||
Db      1  ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGCCTTCCTGGCC 60
QY      21  LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db      |||
Db      61  CTCGTTTCTCGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGAGCGCCT 120
QY      41  ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluPro 60
Db      |||
Db      121  ACCATGGGCTGGCTGCACCTGGAGCGCTTCATGTGCACACCTTGACTGCCAGGAGAGCCA 180
QY      61  AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db      |||
Db      181  GATTCTCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
```

```

QY      81  TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db      |||
Db      241  TGGAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTGTGGATGCTCCCAAGA 300
QY      101  AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisIleArgGlnLeu 120
Db      |||
Db      301  GATTGAGAAGGAGAGCTTCAGGCAGACCCCTCAGCGCTTCTCCTCATGGGATTCGCCAGCTA 360
QY      121  AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAlaAspValGlyAsnLys 140
Db      |||
Db      361  GCTAATATATGTTACACAGCAAGAGCTGAAGCTAGGATTTATGCAGATGTTGGAAATAAA 420
QY      141  ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db      |||
Db      421  ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTGATGCCAGACCTTTGCT 480
QY      161  AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db      |||
Db      481  GACTGGGAGTAGATCTGCTAAAAATTTGATGGTTGTACTGTGACAGTTTGGAAAAATTG 540
QY      181  AlaAspGlyTyrIleHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db      |||
Db      541  CGAGATGGTTATAAGCACATGCTCTGGCCCTGAATAGAGCTGGCAGAGCATTTGTGTAC 600
QY      201  SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db      |||
Db      601  TCCTGTGAGTGGCTCTTTATATGTGCCCTTTCAAAAGCCCAATTTATCAGAAATCCGA 660
QY      221  GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db      |||
Db      661  CAGTACTGCAATCAGCTGGGCAAAATTTTGTGCAATTTGATGATTCCTGGAAAAAGTATAAAG 720
QY      241  SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db      |||
Db      721  AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTTGATGTTGCTGCACAGGG 780
QY      261  GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db      |||
Db      781  GGTGTGAATGACCCAGATATGTTAGTGTGCACTTTGGCCTCAGCTGGAATCAGCAA 840
QY      281  ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db      |||
Db      841  GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTATGCTTAATGACCTC 900
QY      301  ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db      |||
Db      901  CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGATGAAGCGCTAATTCGCAATCAAT 960
QY      321  GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db      |||
Db      961  CAGGACCCCTTGGGCAAGCAAGGTACCGCTTAGACAGGAGCAACTTTGAAGTGG 1020
QY      341  GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db      |||
Db      1021  GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCCGCGAGAGATTGGT 1080
QY      361  GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      |||
Db      1081  GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGAGTGGCTGTATCCT 1140
QY      381  AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db      |||
Db      1141  GCCTGCTTCATCACACAGCTTCCCTGTGAAAGGAAGCTAGGGTTCTATGATGGACT 1200
QY      401  SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db      |||
Db      1201  TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTA 1251

RESULT 5
AR653699
LOCUS      1272 bp      DNA      linear      PAT 13-JUN-2005
DEFINITION      Sequence 17 from patent US 6887696.
ACCESSION      AR653699
```

```
VERSION AR653699.1 GI:67584284
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1272)
AUTHORS Garger, S.J., Turpen, T.H. and Kumagai, M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 17 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source Location/Qualifiers
1..1272
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 1,31e-211 Length: 1272
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-602-220-16 (1-417) x AR653699 (1-1272)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTCGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCGGACATCCCTGGGGCTAGAGCACTGGCAATGATGGCAAGACGCCCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluPro 60
Db 121 ACCATGGGCTGGCTGGCACTGGAGCGCTTCATGGCAACCTTGACTCCAGGAGAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpIleAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGATCAGGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGCAGACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATGTTTACAGCAAGGAGCTGAAGCTAGGATTTATGACATGTTGGAAATATA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATACCTACGACATGATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGACTGCTAAATTTGATGGTTGTACTGTGACATTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGTTTATAGACACATGCTCTTGGCCCTGAAATAGGACTGGCAGAACCATTTGT 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGGCTCTTTATATATGTCGCCCTTTCAAAGGCCCAATTATACAGAAATCC 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGGCTGACATGATGATTCCTGGAAAGATATAAG 720
```

```
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGACTGGACATCTTTTAAACAGAGAGATTTGTATGTTGTGGACCAAGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGATGACCAAGATATGTTAGTATGGCACTTTGGCCCTCAGCTGGATCAGCA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAspLeu 300
Db 841 GTAACTCAGATGCCCTCTCGGCTATCATGGCTGCTCTTTATTTCATGTCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCA 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACCACTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGATGATGATAACCGCAGGAGATTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCGCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValIysArgLysGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCCCTCCCTGTGAAAGGAAGCTAGGTTCTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAGGTTAAGAGTACATATAATCCACAGGACACTGTTTGTCTTCAGCTA 1251
RESULT 6
LOCUS AR654967 1272 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 17 from patent US 6890748.
ACCESSION AR654967
VERSION AR654967.1 GI:67586506
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1272)
AUTHORS Garger, S.J., Turpen, T.H. and Kumagai, M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 17 10-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source Location/Qualifiers
1..1272
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 1,31e-211 Length: 1272
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-602-220-16 (1-417) x AR654967 (1-1272)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTCGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCGGACATCCCTGGGGCTAGAGCACTGGCAATGATGGCAAGACGCCCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluPro 60
Db 121 ACCATGGGCTGGCTGGCACTGGAGCGCTTCATGGCAACCTTGACTCCAGGAGAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpIleAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGATCAGGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGCAGACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATGTTTACAGCAAGGAGCTGAAGCTAGGATTTATGACATGTTGGAAATATA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATACCTACGACATGATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGACTGCTAAATTTGATGGTTGTACTGTGACATTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGTTTATAGACACATGCTCTTGGCCCTGAAATAGGACTGGCAGAACCATTTGT 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGGCTCTTTATATATGTCGCCCTTTCAAAGGCCCAATTATACAGAAATCC 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGGCTGACATGATGATTCCTGGAAAGATATAAG 720
```

Db 61 CTCTGTTCTCGGACATCCCTGGGGCTAGAGCACTGGACATGATGGCAAGACGCCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGTCATAGTACAGAGCTTCCTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysAlaAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGCAGGTATAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTCCTCATGGATTGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATATGTTCCACAGCAAGACTCAAGCTAGGATTTATGCAGATGTTGGAAATAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGAGGCTTCCTGGAGTGTGGTACTACGACATGATGTCGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGACTCTGCTGAAATTTGATGGTGTGGTGTGGAAATTTG 540
Qy 181 AlaAspGlyTyrIlyshMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATTAAGCACATGCTCTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTATC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTCAAAGGCCCAATTAACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGCTGCAATTGATGATTCCTGGAAAGATATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnLysArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGACTGGACATCTTTAAACCAGAGAGAAATGTTGATGTGCTGGACACGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db 781 GGTGGNATGACCCAGATATGTTAGTATGGCAACTTTGGCCTCAGCTGGATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGCCCTCTGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGACGTAATTTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGCAAGCAAGGGTACCACTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAAGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATAAACCGCAGGAGATTGCT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGAACCTCGCTCTATACCATCGCAGTTGCTTCCTGGTAAAGAGTGGCTGTAACTCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400

Db 1141 GCCTGCTTCATCACAGCTCCTCCCTGTGTAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTACATAAATCCACAGGCACCTGTTTGTCTCAGCTA 1251
RESULT 7
AR653694 1278 bp DNA linear PAT 13-JUN-2005
LOCUS Sequence 7 from patent US 6887696.
DEFINITION AR653694
ACCESSION AR653694
VERSION AR653694.1 GI:67584279
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1278)
AUTHORS Gargner,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 7 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source 1..1278
Location/Qualifiers
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 1.32e-211 Length: 1278
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-602-220-16 (1-417) x AR653694 (1-1278)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTCCGCTTCCCTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCTGTTCTCTGGACATCCCTGGGGCTAGAGCACTGGACATGATGATGGCAAGACGCCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGTCATAGTACAGAGCTTCCTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysAlaAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGCAGGTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTCCTCATGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATATGTTCCACAGCAAGACTCAAGCTAGGATTTATGCAGATGTTGGAAATAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGAGGCTTCCTGGAGTGTGGTACTACGACATGATGTCGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGACTCTGCTGAAATTTGATGGTGTGGTGTGGAAATTTG 540

Qy 181 AlaAspGlyTyrIysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 Db 541 GCAGATGGTTATAGACACATGCTCCCTGGCCCTGAATAGGACTGGCAGAGCAATGTGTAC 600
 Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnIysProAsnTyrThrGluIleArg 220
 Db 601 TCCCTGTGAGTGGCCCTCTTTATATATGTGGCCCTTTCAAAGGCCAATATATACAGAAATCCGA 660
 Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspSerTrpIysSerIleLys 240
 Db 661 CAGTACTGCAATCACTGCGCAATTTTGTGTGACATTCATGATTCCTGGAAAAGTATAAAG 720
 Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 Db 721 AGTATCTTGGACTGGACATCTTTTAAACCAGGAGAGAAATGTGTATGTGTGGACACAGG 780
 Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
 Db 781 GGTGGAAATGACCCAGATATGTAGTGTATGGCACTTTGGCCCTCAGCTGGAATCAGCAA 840
 Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
 Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCCAGGATAAGGACGTAATGGCCATCAAT 960
 Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGAGACAATTTGAAGTGTGG 1020
 Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 Db 1021 GAACGACTCTCTCAGCGTTAGCTGGCTGTAGCTATGATAACCGCAGGAGATGGT 1080
 Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCTGTAAATCCT 1140
 Qy 381 AlaCysPheIleThrGlnLeuLeuProValIysArgIysLeuGlyPheTyrGluTrpThr 400
 Db 1141 GCCTGTCTTCATCACACAGCTCCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
 Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
 Db 1201 TCAAGGTTAGAGTCACTAATATCCACAGGACACTGTTTGTCTCAGCTA 1251

RESULT 8
 AR654962
 LOCUS
 DEFINITION Sequence 7 from patent US 6890748.
 ACCESSION AR654962
 VERSION AR654962.1 GI:67586501
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 1278)
 AUTHORS Garger,S.J., turpen,T.H. and Kunagai,M.H.
 TITLE Production of lysosomal enzymes in plants by transient expression
 JOURNAL Patent: US 6890748-A 7 10-MAY-2005;
 Large Scale Biology Corporation; Vacaville, CA
 FEATURES
 source Location/Qualifiers
 1..1278
 /organism="unknown"
 /mol_type="genomic DNA"

ORIGIN

Alignment Scores:
 Pred. No.: 1.32e-211 Length: 1278
 Score: 2293.00 Matches: 417
 Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0
 US-10-602-220-16 (1-417) x AR654962 (1-1278)
 Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCGCTTCCTCGGCC 60
 Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 Db 61 CTGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAAATGATGTGGCAGACGCGCT 120
 Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 Db 121 ACCATGGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCA 180
 Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db 181 GATTCTCTGCATCAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
 Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 241 TGGAGAGATGCAGGTATGAGTACCTCTGCATTTGATGACTGTGGATGGCTCCCAAGA 300
 Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 Db 301 GATTCAAGAGGAGACTTCAGGCAGACCCCTCAGCGCTTTCTCATGGATTGGCAGACTA 360
 Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 Db 361 GCTAATATATGTTCAACGCAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAAATAAA 420
 Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
 Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTCATGCCAGACCTTTGCT 480
 Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 Db 481 GACTGGGAGTAGACTGCTGCTAATTTGATGGTTGTACTGTGACAGTTTGGAAATTTG 540
 Qy 181 AlaAspGlyTyrIysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 Db 541 GCAGATGGTTATAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATGTGTAC 600
 Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnIysProAsnTyrThrGluIleArg 220
 Db 601 TCTGTGAGTGGCCCTCTTTATATGTGGCCCTTTCAAAGGCCAATATATACAGAAATCCGA 660
 Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspSerTrpIysSerIleLys 240
 Db 661 CAGTACTGCAATCACTGGCGAATTTTGTGTGACATTCATGATTCCTGGAAAAGTATAAAG 720
 Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 Db 721 AGTATCTTGGACTGGACATCTTTTAAACCAGGAGAGAAATTTGTATGTGTGGACAGGG 780
 Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
 Db 781 GGTGGAAATGACCCAGATATGTAGTGTATGGCACTTTGGCCCTCAGCTGGAATCAGCAA 840
 Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
 Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCCAGGATAAGGACGTAATGGCCATCAAT 960
 Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGAGACAATTTGAAGTGTGG 1020

```
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGCTTAGCTGGCTGTAGCTATGATATAACCGCAGGAGATGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGCACCTCGCTCTATACCATCGCAGTGTCTTCCCTGGGTAAAGAGTGGCTGTATATCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuLeuLeu 417
Db 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTGTCTTACAGCTA 1251

RESULT 9
AR653697 1284 bp DNA linear PAT 13-JUN-2005
LOCUS
DEFINITION Sequence 13 from patent US 6887696.
ACCESSION AR653697
VERSION AR653697.1 GI:67584282
KEYWORDS
SOURCE
ORGANISM
Unclassified.
REFERENCE
1 (bases 1 to 1284)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 13 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source
1..1284
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 1,33e-211 Length: 1284
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-16 (1-417) x AR653697 (1-1284)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACCTACATCTGGGCTGCGCGCTTGGCGTTCTCGCTTCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGCTTTCCTGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACCTGGAGCGCTTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTTCCTGCAATCAGTGAGAGACTCTTCATGGAGNTGGCAGAGCTCATGGTCTCAGAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGAAGGATGCAGGTATATGATACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGCAGACTTCAGGCAGACCTCAGCGCTTCTCATGGATTCGCACAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
```

```
Db 361 GCTAATATATGTTTACACAGCAAGGACTGAAGACTAGGGAATTTATGCAGATGTGGAAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTTGGATACCTACGACATGTATGATGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTG 540
QY 181 AlaAspGlyTyrIlyHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAAGCATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTCTGACATTTGATGATTCCTTGGAAAAGTATAAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGACCTGGACATCTTTTAAACAGGAGAGAAATTTGTTGATGTTGCTGGACCAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGTGAATGACCCAGATATGTTAGTGATTTGGCACTTTGGCCCTCAGCTGGAAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACCACTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTTGGCTGTAGCTATGATAAACCGCAGGAGATGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGCACCTCGCTCTTATACCATCGCAGTGTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuLeuLeu 417
Db 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTGTCTTACAGCTA 1251

RESULT 10
AR654965 1284 bp DNA linear PAT 13-JUN-2005
LOCUS
DEFINITION Sequence 13 from patent US 6890748.
ACCESSION AR654965
VERSION AR654965.1 GI:67586504
KEYWORDS
SOURCE
ORGANISM
Unknown.
Unclassified.
REFERENCE
1 (bases 1 to 1284)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 13 10-MAY-2005;
```

Large Scale Biology Corporation; Vacaville, CA

FEATURES

Location/Qualifiers
1..1284
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:

Pred. No.: 1.33e-211 Length: 1284
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-16 (1-417) x AR654965 (1-1284)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCTCGTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGGTTCTTGGGACATCCCTGGGGCTAGAGCACTGGACATGGATTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
Qy 61 AspSerCysIleSerGluLeuLeuPheMetGluMetAlaLeuLeuMetValSerGluGly 80
Db 181 GATTCTGTCATCAGTGAAGAGCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLeuAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGCGAGTTATGATGACTCTGATGATGACTGTGGTGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAAGAGGACACTTCAGGAGAGCCCTCAGCGCTTCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTrpValHisSerLeuGlyLeuLeuLeuGlyIleTrpAlaAspValGlyAsnLeu 140
Db 361 GCTAATTTATGTTTACAGCAAGGACTGAAGCTAGGATTTATGAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATGATGACATGATGATGATGATGATGATGAT 480
Qy 161 AspTrpGlyValAspLeuLeuLeuPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAATTTG 540
Qy 181 AlaAspGlyTrpLeuHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
Db 541 GCAGATGGTTATAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAACATTTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
Db 601 TCTGTGGTGGCTCTTTATATGTGGCCCTTTCAAAGGCCCAATTATACAGAAATCCGA 660
Qy 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGATGATGATGATGATGATGATGATGATGAT 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGACATGGACATCTTTTAAACAGAGAGAAATTTGTGATGTTGTGAGACAGGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGATGACCAAGATATGTTAGTGATTGGCAACTTTGGCTTCAGCTGGAAATCAGCAA 840

Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTCTAATGACTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGTAAGGACGTAATTCCTCAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGCAAGCAAGGTTACCACTGTAGACAGGAGACAACTTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGCTTTAGCTGGGCTGTAGCTATGATAAACCCGCGAGGATTTGGT 1080
Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGTAAAGAGTGGCTGTATTCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGACTCTCTCCCTGTGAAAGGAAGCTAGGTTCTATGATGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTTGCTTCAGCTA 1251

RESULT 11

CS135670
LOCUS CS135670 1290 bp DNA linear PAT 09-AUG-2005
DEFINITION Sequence 1 from Patent WO2005069192.
ACCESSION CS135670
VERSION CS135670.1 GI:72056330

KEYWORDS

Source
Homo sapiens (human)
Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.

REFERENCE

1
Garman, S.C.
Crystal structure of human a-galactosidase
Patent: WO 2005069192-A 1 28-JUL-2005;
Transkaryotic Therapies, Inc. (US); Selden, Richard F. (US);
Garboczi, David N. (US); Treco, Douglas A. (US); NIH/NIH (US);
Garman, Scott C. (US); Borowski, Marianne (US); Kinoshita, Carol M.
(US)

FEATURES

Location/Qualifiers

source

1..1290
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.: 1.34e-211 Length: 1290
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-16 (1-417) x CS135670 (1-1290)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCTCGTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGGTTCTTGGGACATCCCTGGGCTAGAGCACTGGACATGGATTGGCAAGGACGCT 120


```
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCAGTGGAGCGCTTCATGTGCAACTTGACTGCCAGGAGGCCA 180

QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTCAGAGGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAGGCC 240

QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGCAGGTATAGTAGTACCTCGCATTTGATGACTGTGTGGATGGCTCCCAAAGA 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGCGAGACTTCAGGCGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATGTTTACACAGCAAGGACTGAAGCTAGGAGTTATGCAGATGTGGAAATAAA 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGGCTTCCTGGGAGTTTTGGATACTACGACATTGATGCCAGACCTTTGCT 480

QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGATCTGCTAATAATTTGATGGTTGTACTGTGCAGATTTGGAAATTTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATGTGTAC 600

QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGGCTCTTTATATATGTGGCCCTTTCAAAGGCCCAATTTATACAGAAATCCGA 660

QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCNATCCTGGCGAAATTTTGTCTGACATTTGATGATTCCTGGAAAGTATTAAG 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGACTGGACATCTTTTAACACGAGAGAGAAATTTGATGTGTGGCAGCAGGG 780

QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGTGAATGACCCAGCATATGTAGTGAATGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840

QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCATATGACCTC 900

QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGACGCTAATTTGCCATCAAT 960

QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnClyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGCGAAGCAAGGTTACCACTGTAGACAGGAGCAACTTTGAAGTGTGG 1020

QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATATAAACCGGACAGGAGATTGGT 1080

QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTGGGTAAGAGAGTGGCTGTAACTCT 1140

QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGATGGACT 1200

QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
```

```
Db 1201 TCAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTGTCTCAGCTA 1251
```

RESULT 12

```
AR653692
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE
ORIGIN
```

Sequence 3 from patent US 6887696.
AR653692
AR653692.1 GI:67584277
Unknown.
Unknown.
Unclassified.
1 (bases 1 to 1290)
Garger, S.J., Turpen, T.H. and Kumagai, M.H.
Production of lysosomal enzymes in plants by transient expression
Patent: US 6887696-A 3 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
Location/Qualifiers
1..1290
/organism="unknown"
/mol_type="genomic DNA"

Alignment Scores:

```
Pred. No.: 1,34e-211 Length: 1290
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
```

US-10-602-220-16 (1-417) x AR653692 (1-1290)

```
QY 1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTTACATCTGGGCTGGCGCTTTCGCTTCCTCGTGGCC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACATGGATGGCAAGGACGCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCAGTGGAGCGCTTCATGTGCAACTTGACTGCCAGGAGGCCA 180

QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTCAGAGGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAGGCC 240

QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGCAGGTATAGTAGTACCTTCGATTTGATGACTGTGTGGATGGCTCCCAAAGA 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGCGAGACTTCAGGCGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATGTTTACACAGCAAGGACTGAAGCTAGGAGTTTATGCAGATGTGGAAATAAA 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGGCTTCCTGGGAGTTTTGGATACTACGACATTGATGCCAGACCTTTGCT 480

QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGATCTGCTAATAATTTGATGGTTGTACTGTGCAGATTTGGAAATTTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATGTGTAC 600
```



```
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
|
|
|
Db 601 TCCTGTGAGTGGCCCTCTTTATATATGTGGCCCTTTCAAAAGGCCAATTATACAGAAATCCGA 660
|
|
|
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
|
|
|
Db 661 CAGTACTGCAATCAGTGGCAAAATTTTGTGACATTTGATGATTCCTGGAAAAGTATAAAG 720
|
|
|
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
|
|
|
Db 721 AGTATCTTGACTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTGGACCAAGG 780
|
|
|
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
|
|
|
Db 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
|
|
|
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
|
|
|
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTATGCTTAATGACCTC 900
|
|
|
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
|
|
|
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATTTGCCATCAAT 960
|
|
|
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
|
|
|
Db 961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
|
|
|
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
|
|
|
Db 1021 GAACGACCTCTCTCAGGCTTAGCGCTGGCTGTAGCTATGATGAACCCGGCAGAGATGGT 1080
|
|
|
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
|
|
|
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGAGTGGCTGTATCCT 1140
|
|
|
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
|
|
|
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
|
|
|
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
|
|
|
Db 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTA 1251
|
|
|
RESULT 13
AR654960
LOCUS AR654960 1290 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 3 from patent US 6890748.
ACCESSION AR654960
VERSION AR654960.1 GI:67586497
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1290)
AUTHORS Garger, S. J., Turpen, T. H. and Kumagai, M. H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 3 10-MAY-2005.
FEATURES
Location/Qualifiers
Source 1..1290
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 1,34e-211 Length: 1290
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
```

```
US-10-602-220-16 (1-417) x AR654960 (1-1290)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
|
|
|
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCGGCC 60
|
|
|
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
|
|
|
Db 61 CTCGTTTCTCTGGGACATCCCTGGGGCTAGAGCACTGGACCAATGATTTGGCAAGACGCT 120
|
|
|
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluLeuPro 60
|
|
|
Db 121 ACCATGGGCTGGCTGGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGAAAGGCCA 180
|
|
|
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
|
|
|
Db 181 GATTCTCGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
|
|
|
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
|
|
|
Db 241 TGGAAAGGATGCAAGTTATGAGTACCTCTGCAATTCATGACTGTTGGATGGCTCCCCAAGA 300
|
|
|
QY 101 AspSerGluGlyValArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
|
|
|
Db 301 GATTCAAGAGGACAGACTTCAGGAGACCCCTCAGCGCTTCTCATGGGATTCGCCAGCTA 360
|
|
|
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
|
|
|
Db 361 GCTAATTTATGTTACAGCAAGGACGCTGAAGCTAGGAGTTTATGCAGATGTTGGAAATAA 420
|
|
|
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
|
|
|
Db 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATACCTACGACATTCATGATGCCAGACCTTTGCT 480
|
|
|
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
|
|
|
Db 481 GACTGGGGAGTAGATCTGTAAATTTTGGTGGTGTGTACTGTGACAGTTTGGAAAATTTG 540
|
|
|
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
|
|
|
Db 541 GCAGATGGTTATTAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGAC 600
|
|
|
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
|
|
|
Db 601 TCCTGTGAGTGGCCCTCTTTATATGTGGCCCTTTCAAAAGGCCAATTATACAGAAATCCGA 660
|
|
|
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
|
|
|
Db 661 CAGTACTGCAATCAGTGGGCAAAATTTTGTGACATTTGATGATTCCTGGAAAAGTATAAAG 720
|
|
|
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
|
|
|
Db 721 AGTATCTTGACTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTGGACCAAGG 780
|
|
|
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
|
|
|
Db 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
|
|
|
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
|
|
|
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTATGCTTAATGACCTC 900
|
|
|
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
|
|
|
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATTTGCCATCAAT 960
|
|
|
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
|
|
|
Db 961 CAGGACCCCTTGGGCAAGCAAGGTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
|
|
|
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
|
|
|
```

```

Db      1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAAACCGCAGCAGATTGCT 1080

Qy      361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyIysGlyValAlaCysAsnPro 380
|||||
Db      1081 GGACCTCGCTCTATACCATCGAGTGTCTCCCTGGGTAAAGGAGTGGCCCTGTATCTCT 1140

Qy      381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
|||||
Db      1141 GCTGCTTCTATCACAGCTCTCTCCCTGTGAAAGGAAGCTAGGTTCTATGATGGACT 1200

Qy      401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
|||||
Db      1201 TCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTGTCTCAGCTA 1251

RESULT 14
BT006864
LOCUS      1290 bp mRNA linear PRI 13-MAY-2003
DEFINITION Homo sapiens galactosidase, alpha mRNA, complete cds.
ACCESSION BT006864
VERSION BT006864.1 GI:30582566
KEYWORDS FLI_CDNA.
SOURCE Homo sapiens
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
            1 (bases 1 to 1290)
            Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
            Cloning of human full-length cDNAs in BD Creator(TM) System Donor
            vector
            Unpublished
            2 (bases 1 to 1290)
            Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
            Direct Submission
            Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
            Circle, Palo Alto, CA 94303, USA
            This cDNA clone is a part of a collection of human full length
            expression clones generated by BD Biosciences Clontech and the
            Harvard Institute of Proteomics. Each cDNA has been cloned in two
            forms: with and without stop-codon (to allow fusion with C-terminal
            tag). The cDNA has been directionally cloned using BD In-Fusion(TM)
            cloning system between the SalI and HindIII sites of the pDNR-DUAL
            vector. Additional sequences in the clone: 'ACC' after SalI site
            and before 'ATG' to provide Kozak consensus sequence; 'GG' after
            last codon and before HindIII site to maintain reading frame.
            Clone distribution: http://bioinfo.clontech.com/orfclones.
            Location/Qualifiers
                1..1290
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone_lib="BD Creator(TM) CDS Library derived from MGC
                    collection"
                    /lab_host="DH5alpha T1 resistant"
                    /note="Vector: pDNR-Dual"
                1..1290
                    /codon_start=1
                    /product="galactosidase, alpha"
                    /protein_id="AAP35510.1"
                    /db_xref="GI:30582567"
                    /translation="MQLRNPGLHGCALRFLALVSWDIPGARALDNLGLARTPTMGW
                    LWERPMCNLDQEEPDSCISEKLFMEAEALMVSEGWKDGAYEYLICIDCWMAPOKDS
                    ERLQADPQRFPHGIRQLANYHSHKGLGIYADVGNKTCAGFPSCGXYDIDACQFPA
                    DWGVDLLKDFGDCSLDENLADYGHKMSLALNKTGRISVTSCWFLPMFPFKPNYTE
                    INQYCNHWRNALIDDSWSKISILDTSFQNERIVDVAGPGWNPDPMLVIGNFGLS
                    WNQVQTOMALWAIAPLEFMSNLDLHISFPQAKALLQDKVIADINQDPLGKQGYQLPVR
                    DNFVWVRPLSLAWAVAMINROEIGGPRSYTTAVASLGKGVACNPACFITQLLPVKR
                    KLGFEYWTSLRSHINPTGTVLLQLENTMQMSLKDLL"
FEATURES
    source
    CDS

```

ORIGIN

```

Alignment Scores:      1.34e-211      Length:      1290
Pred. No.:      2293.00      Matches:      417
Score:      100.00%      Conservative: 0
Percent Similarity:      100.00%      Mismatches: 0
Best Local Similarity:      100.00%      Indels: 0
Query Match:      100.00%      Gaps: 0
DB:
US-10-602-220-16 (1-417) x BT006864 (1-1290)

Qy      1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
|||||
Db      1 ATGCAGCTGAGGACCCAGAACTACATCTGGGCTGGCGCTTGGCGCTTCGCTTCCTGGCC 60

Qy      21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
|||||
Db      61 CTCTGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAGAGCAGCCT 120

Qy      41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
|||||
Db      121 ACCATGGGGCTGGCTGCACCTGGGAGCGCTTTCATGTGCACCTTGACTGCCAGGAGAGCCA 180

Qy      61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
|||||
Db      181 GATTCTCTGCATCAGTGAGAGGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

Qy      81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
|||||
Db      241 TGGAGGATGCAGGTATTGAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCAAGA 300

Qy      101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
|||||
Db      301 GATTTCAGAGGAGCAGACTTCAGGCAGACCTTCAGCGCTTCTCTCATGGATTCCCAAGCTA 360

Qy      121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
|||||
Db      361 GCTAATTATGTTTCACAGCAAGGAGCTGAAGCTAGGATTTATGCAGATGTTGGAATAAA 420

Qy      141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspIleAspIleAspIle 160
|||||
Db      421 ACCTGCGCAGGCTCTCCCTGGGAGTTTTGGATATCTACGACATTGATGCCAGACCTTTGCT 480

Qy      161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
|||||
Db      481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTTACTGTGACAGTTTGGAAATTTG 540

Qy      181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
|||||
Db      541 GCAGATGGTTATAAGCACATGCTTGGCCCTGAATAGGACTGGCAGAGCAATTGTGTAC 600

Qy      201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysPheProAsnTyrThrGluIleArg 220
|||||
Db      601 TCCTGTGAGTGGGCTCTTTATATATGTGGCCCTTTCAAAGGCCCAATTATACAGAAATCCGA 660

Qy      221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
|||||
Db      661 CAGTACTGCAATCAGCTGGCGAAATTTTGTGCAATTGATGATTCCTGGAAAGATATAAG 720

Qy      241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
|||||
Db      721 AGTATCTTGGACTGGACATCTTTTAACCAAGGAGAGAAATTTGTGATGTTGTGACCAAGG 780

Qy      261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
|||||
Db      781 GGTTCGAATGACCCAGATATGTTAGTGTAGTTGGCAACTTTGGCCCTCAGCTCGAATTCAGCA 840

Qy      281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
|||||
Db      841 GTAACCTCAGATGGCCCTCTGGGGCTATCATGGCTGCTCTTATTTCATGCTTAATGACCTC 900

Qy      301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
|||||

```

Db 901 CGACATCAGCCCTCAAGCAAGCTCTCTTCCAGGATAAGGACGTAATTCGCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrr 340
Db 961 CAGGACCCCTTGGCAAGCAAGGTACACGCTTAGACAGGAGCAACCTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrrAlaValAlaMetIleAsnArgGlnGluLeGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCGCTGGCTGTAGCTATGATGATAAACCGGAGGAGATTGGT 1080
Qy 361 GlyProArgSerTrrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGAGTTCCTTCCGTGGTAAAGAGTGGCTGTAAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrrThr 400
Db 1141 GCCTGCTTCATCACAGACCTCTCCCTGTGTAAGGAGGCTAGGTTCTATGATGAGCT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTTCAAGCTA 1251
RESULT 15
AY889631
LOCUS
DEFINITION
AY889631 1290 bp mRNA linear SYN 29-MAR-2005
Synthetic construct Homo sapiens clone FLH025769.01X galactosidase
alpha (GLA) mRNA, complete cds.
ACCESSION
AY889631.1 GI:60655716
VERSION
Human ORF Project.
KEYWORDS
synthetic construct
SOURCE
synthetic construct
ORGANISM
other sequences; artificial sequences.
REFERENCE
1 (bases 1 to 1290)
Hines, L., Rolfs, A., Jepson, D., Moreira, D., Raphael, J., Kelley, F.,
Shen, B., Halleck, A., Koundinya, M., Hu, Y., Zuo, D., Taycher, E.,
Williamson, J. and Labaer, J.
Cloning of human full-length CDS in Creator (TM) recombinational
vector system
Unpublished
JOURNAL
2 (bases 1 to 1290)
Hines, L., Rolfs, A., Jepson, D., Moreira, D., Raphael, J., Kelley, F.,
Shen, B., Halleck, A., Koundinya, M., Hu, Y., Zuo, D., Taycher, E.,
Williamson, J. and Labaer, J.
Direct Submission
Submitted (04-JAN-2005) Biological Chemistry and Molecular
Pharmacology, Harvard Institute of Proteomics, 320 Charles St.,
Cambridge, MA 02141, USA
COMMENT
This CDS clone is a part of a collection of human full-length
expression clones generated by Harvard Institute of Proteomics.
This ORF clone has been cloned with normalized stop-codon. The CDS
has been directionally cloned using BD In-Fusion(TM) cloning system
between the SalI and HindIII sites of the pDNR-Dual vector.
Additional sequences in the clone: 'ACC' after SalI site and
before 'ATG' to provide Kozak consensus sequence. Each clone is
clonally isolated and full-length sequence-verified.
FEATURES
Location/Qualifiers
1..1290
/organism="synthetic construct"
/mol_type="mRNA"
/db_xref="taxon:32630"
/clone="FLH025769.01X"
/lab_host="Escherichia coli DH5alpha T1 resistant"
/note="derived from MGC template"
1..1290
/gene="GLA"
1..1290
/gene="GLA"
/codon_start=1
/transl_table=11
/product="galactosidase alpha"
/protein_id="AAK32422.1"

/db_xref="GI:60655717"
/translation="MQLRNPEHLGCALALFLALVSWDIPGARALDNLGARTPTMGW
LHBERFMCNDQBEEDSCISEKLFMEALGIYADGVNKTGAGPFGFYDIDAQTPA
EGRLQADLPQRFPHGIRQLANYVHSGIKLGIYADGVNKTGAGPFGFYDIDAQTPA
DNGVDLLKDFGCVCDLSLEADGYSKMSLALNRTGRSIVVSCWPLMYPFQKPNYTE
IRQYCNHNFENFADIDDSWKSISKILMTSFPNOERIYDVAGPGGWNPDMLVIGNFGLS
WQOQVTCMALMAIAPLFWNDLHHSIPQAKALLQDKDVIAINQDPLGKQGVLRQS
DNFVWERPLSLGAMAVMINQBEIGGPRSYTIAVASLGGVACNACFATQLLPVVR
KLGFEBWTSRLRSHINPTGTVLLENTQMSLKDLL"
ORIGIN
Alignment Scores:
Pred. No.: 1,34e-211 Length: 1290
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 11 Gaps: 0
US-10-602-220-16 (1-417) x AY889631 (1-1290)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAGACTACATCTGGGCTGGCGCTTCGGCTTCCTCTCGGCC 60
Qy 21 LeuValSerTrrAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGACGCT 120
Qy 41 ThrMetGlyTrrLeuHisTrrGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGGCCA 180
Qy 61 AspSerCysIleSerGlyLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrrLysAspAlaGlyTrrGluTrrLeuCysIleAspAspCysTrrMetAlaProGlnArg 100
Db 241 TGGAGGATGCAAGTTATGAGTACCTCTGCATTGATGACTGTTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgGluAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGAGGAGCTTCAGGAGAGCCCTCAGCGCTTCCTCATGGATTCGCCAGCTA 360
Qy 121 AlaAsnTrrValHisSerLysGlyLeuLysLeuGlyIleTrrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGAGGACTGAAGCTAGGATTTATGCAGATGTTGAAATAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrrTrrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCTCGGAGTTTGGATACACGACATGATGATGCCAGACCTTTGCT 480
Qy 161 AspTrrGlyValAspLeuLysPheAspGlyCysTrrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGACAGTTTGGAAAAATTG 540
Qy 181 AlaAspGlyTrrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGCACATGCTCTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTrrProLeuTrrMetTrrProPheGlnLysProAsnTrrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAAGCCCAATATATACAGAAATCCGA 660
Qy 221 GlnTrrCysAsnHisTrrArgAsnPheAlaAspIleAspAspSerTrrLysSerIleLys 240
Db 661 CAGTACTGCATCAGTGGCGAAATTTTGTGTCATTGATGATTCCTGGAAAAAGTATAAAG 720
Qy 241 SerIleLeuAspTrrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGACTGGACATCTTTTAAACAGGAGAGAAATTTGTTGATGTTGCTGGACCCAGG 780

Qy	261	GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln	280
Db	781	GGTTGGAAATGACCCAGAGATATGTTAGTATGGCAACTTGGCCTCAGCTGGAATCAGCAA	840
Qy	281	ValThrGlnMetAlaLeuTrpAlaIleMetAlaIleProLeuPheMetSerAsnAspLeu	300
Db	841	GTAACCTCAGATGGGCCCTCTGGGCTATCATGGCTGCTCTCTTATTCATGTCATATGACCTC	900
Qy	301	ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn	320
Db	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCGCATCAAT	960
Qy	321	GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp	340
Db	961	CAGGACCCCTTGGGCAAGCAGGAGTCCAGCTTAGCAGGAGACAACCTTGAAGTGTGG	1020
Qy	341	GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly	360
Db	1021	GAA CGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGCAGGAGATTGGT	1080
Qy	361	GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro	380
Db	1081	GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGATGGGCTGTAAATCCT	1140
Qy	381	AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr	400
Db	1141	GCCTGCTTCATCACAACAGCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTCATGAATGGACT	1200
Qy	401	SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu	417
Db	1201	TCAAGGTTTAAAGATGCATATAATCCCAAGGCACTGTTTGGCTTCAGCTA	1251

Search completed: January 1, 2006, 03:35:26
Job time : 9545.23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 23:47:52 ; Search time 21.0897 Seconds
(without alignments)
148.074 Million cell updates/sec

Title: US-10-602-220-16

Perfect score: 2293

Sequence: 1 MQRNPELHGCMALRFLA.....EWTSLRSHNPRTGVLLQL 417

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 57103 seqs, 7488799 residues

Total number of hits satisfying chosen parameters: 57103

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New.*

1: /cgn2_6/prodata/1/pubpa/US08_NEW_PUB.pap.*
2: /cgn2_6/prodata/1/pubpa/US06_NEW_PUB.pap.*
3: /cgn2_6/prodata/1/pubpa/US07_NEW_PUB.pap.*
4: /cgn2_6/prodata/1/pubpa/PCT_NEW_PUB.pap.*
5: /cgn2_6/prodata/1/pubpa/US09_NEW_PUB.pap.*
6: /cgn2_6/prodata/1/pubpa/US10_NEW_PUB.pap.*
7: /cgn2_6/prodata/1/pubpa/US11_NEW_PUB.pap.*
8: /cgn2_6/prodata/1/pubpa/US60_NEW_PUB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	85	3.7	437	7	US-11-073-626-3
2	82	3.6	437	7	US-11-073-626-1
3	81	3.5	310	7	US-11-074-176-108
4	81	3.5	944	7	US-11-057-058-68
5	80.5	3.5	414	6	US-10-467-657-32
6	80.5	3.5	414	6	US-10-467-657-8372
7	80	3.5	861	7	US-11-038-284-36
8	79.5	3.5	1827	7	US-11-057-058-62
9	79	3.4	322	6	US-10-467-657-8554
10	79	3.4	368	7	US-11-129-143-62
11	78.5	3.4	622	6	US-10-510-386-236
12	78	3.4	760	6	US-10-821-234-1141
13	78	3.4	786	6	US-10-467-962B-103
14	77.5	3.4	522	6	US-10-131-826A-450
15	77.5	3.4	903	7	US-11-057-058-65
16	77	3.4	734	6	US-10-652-893-2
17	76.5	3.3	320	6	US-10-467-657-3254
18	76.5	3.3	717	6	US-10-793-626-3022
19	75.5	3.3	275	7	US-11-055-822-2
20	75.5	3.3	843	6	US-10-645-441-8
21	75	3.3	402	6	US-10-650-326B-21
22	75	3.3	402	7	US-11-051-568-29
23	75	3.3	627	6	US-10-873-528-191
24	74.5	3.2	482	6	US-10-793-626-1314
25	74.5	3.2	573	6	US-10-467-657-684

26	74.5	3.2	966	7	US-11-057-058-67	Sequence 67, Appl
27	74	3.2	417	6	US-10-467-657-6612	Sequence 6612, Ap
28	73.5	3.2	932	7	US-11-057-058-59	Sequence 59, Appl
29	73	3.2	514	7	US-11-103-037-3	Sequence 3, Appli
30	73	3.2	734	7	US-11-137-465-65	Sequence 65, Appli
31	72	3.1	515	6	US-10-630-203-6	Sequence 6, Appli
32	72	3.1	842	7	US-11-038-284-38	Sequence 38, Appli
33	72	3.1	847	7	US-11-038-284-42	Sequence 42, Appli
34	71.5	3.1	442	6	US-10-485-788A-516	Sequence 516, App
35	71.5	3.1	316	6	US-10-821-234-1526	Sequence 1526, Ap
36	70.5	3.1	834	6	US-10-928-446A-188	Sequence 188, App
37	70.5	3.1	834	6	US-10-928-446A-192	Sequence 192, App
38	70.5	3.1	834	6	US-10-928-446A-194	Sequence 194, App
39	70.5	3.1	834	6	US-10-928-446A-196	Sequence 196, App
40	70.5	3.1	834	6	US-10-928-446A-198	Sequence 198, App
41	70.5	3.1	834	6	US-10-928-446A-200	Sequence 200, App
42	70.5	3.1	834	6	US-10-928-446A-202	Sequence 202, App
43	70.5	3.1	942	6	US-10-928-446A-190	Sequence 190, App
44	70.5	3.1	947	6	US-10-928-446A-182	Sequence 182, App
45	70.5	3.1	950	6	US-10-928-446A-184	Sequence 184, App

ALIGNMENTS

RESULT 1

US-11-073-626-3
; Sequence 3, Application US/11073626
; Publication No. US20050244926A1
; GENERAL INFORMATION:

; APPLICANT: KUROSAWA, KEIKO
; APPLICANT: HIROKAWA, KOZO
; APPLICANT: KAJIYAMA, NAOKI
; TITLE OF INVENTION: Novel fructosyl peptide oxidase
; FILE REFERENCE: 227590US0
; CURRENT APPLICATION NUMBER: US/11/073,626
; CURRENT FILING DATE: 2005-03-08
; PRIOR APPLICATION NUMBER: US/10/232,655
; PRIOR FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: JP 2001-2666665
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: JP 2001-378151
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: JP 2002-228727
; PRIOR FILING DATE: 2002-08-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3

; LENGTH: 437

; TYPE: PRT

; ORGANISM: Eupenicillium terrenum

US-11-073-626-3

Query Match	3.7%	Score 85;	DB 7;	Length 437;
Best Local Similarity	20.4%;	Pred. No. 0.63;		
Matches	46;	Conservative	29;	Mismatches 71; Indels 80; Gaps 12;
QY	32	LDNLARTPTMGWLHWFMCNLDCEEPDSCISEKLFMEAEIMVSEGMKADAGEYLICI	91	
Db	115	LDAGIGLEKTNVWL-----ESEDILAKAPNFTREQV---KGWKG-----LFCT	155	
QY	92	DDCWMAPOBDSRGLQADPQRPFHGIRQLANYVHSGKLKLG-----IYADVGNKT	141	
Db	156	DGGLAAAK-----AINAIGIFLDKGVKFGGAGTFOQPLFAADG-KT	199	
QY	142	CAGFPDSFGYIDIDAGTAD-----NGVDLLKF-DGC-----YCDSELENADGY	184	
Db	200	CIGLETDTG-----TKYFADKVVLAAGASPTLVLEDCQVSKAWFAHQLTPKEADAY	254	
QY	185	KHMSLALNRTGSRISVYSCWPLYMPPFKPNYTEIRQYCNHMRNFA	230	
Db	255	KNV-----PVVYDGEYGF-----FFEFNEYGVIKVCDEFPGFS	287	

```
RESULT 2
US-11-073-626-1
; Sequence 1, Application US/11073626
; Publication No. US20050244926A1
; GENERAL INFORMATION:
; APPLICANT: KUROKAWA, KEIKO
; APPLICANT: HIROKAWA, NAOKI
; APPLICANT: KAJIYAMA, KOZO
; TITLE OF INVENTION: Novel fructosyl peptide oxidase
; FILE REFERENCE: 227590U80
; CURRENT APPLICATION NUMBER: US/11/073,626
; CURRENT FILING DATE: 2005-03-08
; PRIOR APPLICATION NUMBER: US/10/232,655
; PRIOR FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: JP 2001-266665
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: JP 2001-378151
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: JP 2002-228727
; PRIOR FILING DATE: 2002-08-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Coniochaeta sp.
US-11-073-626-1

Query Match 3.6%; Score 82; DB 7; Length 437;
Best Local Similarity 19.5%; Pred. No. 1.2;
Matches 42; Conservative 31; Mismatches 82; Indels 60; Gaps 10;

QY 33 DNGIARTPTMGWLHWRFCMNCDCQEEPDSCISEKLFMEAEMLVSEGKWDAGYVLCID 92
Db 116 DAGAGLEKTHAWL-----DNEDEILSKMPLLRDQI---QGKAIWSQ---D 156
QY 93 DCWMAQRDSEGRLOADPQRFPHGIRQLANYVHSGKGLG-----IYADVGNKTC 142
Db 157 GGLAAAK-----AINAIGQLKRGVKGFGGAGSKQPLFDDEGT-TC 200
QY 143 AGPFGSGF---YYDIDAQTFADMGVDLLKPDGCGDSENLADGKXMSL-----ALNRTG 195
Db 201 IGVETADGTYKADKVVLAGAMSPITLVLEDDQCCSK----AWYAHQILTPEEAAYKG 256
QY 196 RSIVYSCWFLYMWPFQKPNYTIQYCNHWRNPA 230
Db 257 VPVYNGEFGF----PPEPNEFGVIKVCDEFPGFS 287

RESULT 3
US-11-074-176-108
; Sequence 108, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Peril, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 108
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus

US-11-074-176-108
Query Match 3.5%; Score 81; DB 7; Length 310;
Best Local Similarity 20.9%; Pred. No. 1;
Matches 54; Conservative 32; Mismatches 114; Indels 58; Gaps 10;

QY 56 CQEEPDSCISEKLFMEAEMLVSEGKWDAGYVLCIDDCWMAQRDSEGRLOADPQRFPH 115
Db 10 CYNEESI---PLFYNAVE-KVMQTILDDIYEFWFVNDGSSDKSLEMLQKNDPEHVV 65
QY 116 GIRQLANYVHSGKGLGIYADVGNKTCAGFPGSGFYDDIDAQTFADMGVDLLKPDGCG 175
Db 66 YVSFSRNFGEKAAALYAGLQASIGDIYW-----MDVDLQ-----D 100
QY 176 SLENLADGKXMSLA-----LNRTGRSIVYSCWFLYMWPFQKPNYTI-----RQY 222
Db 101 PPEFLPEMYNIYTKGYDCIGTRRVRDKGEAKFSFLSDMFYKLVNKNISNTEIVSGARDY 160
QY 223 CNHWRNPA-----IDDSWKSISILDWTGF-----NOERIVDVAGPGGWNDDPMLVI 270
Db 161 RMTROMVDVAVLSMPYENRFSKIPSWGVFKTKYLDYHNVER---VAGESDWNWTKLKY 217
QY 271 GNFGLS-WNQVVTQMALW 287
Db 218 AMDGIADFSQAPLNIAVM 235

RESULT 4
US-11-057-058-68
; Sequence 68, Application US/11057058
; Publication No. US20050244400A1
; GENERAL INFORMATION:
; APPLICANT: Lebowitz, Jonathan
; APPLICANT: Maga, John
; TITLE OF INVENTION: ACID ALPHA-GLUCOSIDASE AND FRAGMENTS THEREOF
; FILE REFERENCE: SYM-011
; CURRENT APPLICATION NUMBER: US/11/057,058
; CURRENT FILING DATE: 2005-02-10
; PRIOR APPLICATION NUMBER: US 60/543,812
; PRIOR FILING DATE: 2004-02-10
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 68
; LENGTH: 944
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-057-058-68

Query Match 3.5%; Score 81; DB 7; Length 944;
Best Local Similarity 26.5%; Pred. No. 4.4;
Matches 36; Conservative 13; Mismatches 49; Indels 38; Gaps 7;

QY 75 LMVSEGKWDAGYVLCIDDCWM-APQDSEGRLOADPQRFPHG---IRQLAN----- 122
Db 410 LEVDQGFDD---HNLPCDVIWLDIEHADGKRYFTWDPSPFPQPRTMLERLASKRKLVAI 466
QY 123 -----YVHSGKGLGIYADVGNKTCAG-----FPGSGFYDDIDAQTFADMGV 164
Db 467 VDPHIKVDGSGYRVHEELRNGLYV---KTRDGSDEYGCWPGSAGYPDTNPTMRAWWA 522
QY 165 DLLKFDGCGDSENL 180
Db 523 NMFSDYN-YEGSAPNL 537

RESULT 5
US-10-467-657-32
; Sequence 32, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
```

```

QY 137 VGKNTCAGPFGSGYYDIDAQTADWGVDDLKPKDGCYCDSLENIADGVYKHMSLALNRRTGR 190
Db 368 MVNE-TEGWPHA-----EFDG-FCDLNLCNLIQAEKY-----GK 398

QY 197 SIVYSCWPLY 207
Db 399 KHYFPCSLPLF 409

RESULT 7
US-11-038-284-36
; Sequence 36, Application US/11038284
; Publication No. US20050246793A1
; GENERAL INFORMATION:
; APPLICANT: COOKE, DAVID
; APPLICANT: DEBET, MARTINE
; APPLICANT: GIDLEY, MICHAEL, JOHN
; APPLICANT: JOBLING, STEPHEN, ALAN
; APPLICANT: SAFFORD, RICHARD
; APPLICANT: SIDEBOTTOM, CHRISTOPHER, MICHAEL
; APPLICANT: WESTCOTT, ROGER, JOHN
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO PLANT STARCH COMPOSITION
; FILE REFERENCE: 054163-5003-US
; CURRENT APPLICATION NUMBER: US/11/038,284
; CURRENT FILING DATE: 2005-01-21
; PRIOR APPLICATION NUMBER: US/10/056,454
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: PCT/GB96/01075
; PRIOR FILING DATE: 1996-05-03
; PRIOR APPLICATION NUMBER: GB 9607409.1
; PRIOR FILING DATE: 1996-04-10
; PRIOR APPLICATION NUMBER: GB 9509229.2
; PRIOR FILING DATE: 1995-05-05
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 36
; LENGTH: 861
; TYPE: PRT
; ORGANISM: Lathyrus sp.
US-11-038-284-36

Query Match 3.5%; Score 80; DB 7; Length 861;
Best Local Similarity 25.6%; Pred. No. 4.9;
Matches 32; Conservative 17; Mismatches 42; Indels 34; Gaps 6

QY 108 ADPQRF--PHGIRQLANYHVSKGJLKL---GIYADVGNKTCAGF-----PGSFG 150
Db 395 APSRRGTGTPEDLLKSLIDRAHELGLLVLMDIVHSSNNLTDLGNMFDGTDGHHFPGSRG 454

QY 151 YYDI-DAQTF-----ADMGVDLLKFPDGCYCDSLENIADGVYKHMSLALNRGTG 195
Db 455 YHMWMSRLPNYGSWEVLRVLLSNARWWLDEYKFDGFRFDGVTSMW--YTHHGLQVSTG 512

QY 196 RSIVY 200
Db 513 NYSEY 517

RESULT 8
US-11-057-058-62
; Sequence 62, Application US/11057058
; Publication No. US20050244400A1
; GENERAL INFORMATION:
; APPLICANT: LeBowitz, Jonathan
; APPLICANT: Mega, John
; TITLE OF INVENTION: ACID ALPHA-GLUCOSIDASE AND FRAGMENTS THEREOF
; FILE REFERENCE: SYM-011
; CURRENT APPLICATION NUMBER: US/11/057,058
; CURRENT FILING DATE: 2005-02-10
; PRIOR APPLICATION NUMBER: US 60/543,812
; PRIOR FILING DATE: 2004-02-10
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: Patent in version 3.3

```

```
; SEQ ID NO 62
; LENGTH: 1827
; TYPE: PRT
; ORGANISM: Rabbit sp.
US-11-057-058-62

Query Match      3.5%; Score 79.5; DB 7; Length 1827;
Best Local Similarity 18.3%; Pred. No. 15;
Matches 42; Conservative 29; Mismatches 61; Indels 97; Gaps 10;

QY 143 AGPPSGFYDIDQTFADMGV-----DLKFGDCYCD-----175
DB 1360 AAPP-----DFFRNSTAEWWTREILDFYNNYKFKDGLWIDMNEPSSFVNGTTTNNVCRNT 1413
QY 176 -----SLENADGKXMSLAIN-----RTGRSIVYSCWPLYMWPQKKNYTEIR- 220
DB 1414 ELNYPYPPELTKRTDGLHFTMCMETEHILSDGSSVLHYDVHNLGYMSQAKPTYDALQK 1473
QY 221 -----QYCNHW--RNFADIDDSWKSISILDWTSFNQERIVDVAGP- 259
DB 1474 TTGKRGIVISRSTYPTAGRWAGHGLGNDYARWDMNDSIIGMWFSLFG-----ISYTGAD 1529
QY 260 --CGWNPDMMLV-----IGNF-----GLSWNQOVTQM 284
DB 1530 ICGFFNDSYHLCTRWTLQGFAPFARNHNIQFTRRQDPVSMQTFVEM 1578

RESULT 9
US-10-467-657-8554
; Sequence 8554, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; PRIOR FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: Seqwin99, version 1.04
; SEQ ID NO 8554
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-8554

Query Match      3.4%; Score 79; DB 6; Length 322;
Best Local Similarity 20.7%; Pred. No. 1.7;
Matches 78; Conservative 38; Mismatches 101; Indels 160; Gaps 23;

QY 2 QLNPELHLGCALALRFALVSWDIPCARALDNGCLARTPTMGWLHWFRCNLDCCQEPD 61
DB 27 QCSDELHHLCEAG-----FAGIARLHPFVKVHVWKQWRK-----64
QY 62 SCISEKLFMEALMWSEGWKDAQ-----YEYLCIDDCWMAQPDSEGRLOA---108
DB 65 -----HLFQAEWTREMGRLKQALLQEAQDFVL-----DSQGLKSACF 102
QY 109 -----DPORFPHGIRQLA---NYVHSKGLKL-----GIYADVGNKTCAGPGSP 149
DB 103 AKMAKSPYGLDKHSAREGVAALAYDKYVVPKGRNAVWRNRLFAQV-----F 151
QY 150 GYDIDAQTFADMGVLLKFDGCGYCDSELENADGKXMSLAINRTGRSIVYSCWPLYM 209
DB 152 GYAMPETQVF---GLSVPE-----ASRLKNLAQPY---YALHATSD---SKLWPMENW 197
QY 210 P--FQKNPNYTEIROCYCN---HWRNFADIDDSWKSISILDWTSFNQERIVDVAGPGGWN 264
```

```
Db 198 RELLOKLNQ---KQCNNVYLPWGNFAE-----KVRAEQIAD-----230
QY 265 PDMVLVGNFGLSWNQOVTQOMALWMAAPLF-----MSNDLRHISPOAKALLODKDV 316
Db 231 -----GLPFTIIVCAKMWNL--LQAYLLKHAIVGIVGDTGLLHL---ANAL--EKPV 274
QY 317 IAI--NQDPLGKQGYQL 331
Db 275 VGIYTDTDPI-KTGVQV 290

RESULT 10
US-11-129-143-62
; Sequence 62, Application US/11129143
; Publication No. US20050266518A1
; GENERAL INFORMATION:
; APPLICANT: BERRY, Alan
; APPLICANT: BRETZEL, Werner
; APPLICANT: HUMBELIN, Markus
; APPLICANT: LOPEZ-ULIBARRI, Rual
; APPLICANT: MAYER, Anne F.
; APPLICANT: YELISEEV, Alexei A.
; TITLE OF INVENTION: IMPROVED ISOPRENOID PRODUCTION
; FILE REFERENCE: C38435/121966
; CURRENT APPLICATION NUMBER: US/11/129,143
; CURRENT FILING DATE: 2005-05-13
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Sulfolobus solfataricus
US-11-129-143-62

Query Match      3.4%; Score 79; DB 7; Length 368;
Best Local Similarity 25.6%; Pred. No. 2;
Matches 50; Conservative 32; Mismatches 79; Indels 34; Gaps 11;

QY 176 SLENADGKXMS--LALNRTGRSIVYSCWPLYMWPQKPNYTEIROCYCNHWRNPADID 233
DB 175 ALERLRDISKELSVPIIVKESNGI--SMETAKLISYGIKNFTDSGGGTNNIAIEMIR 232
QY 234 D-----SWK--SIKSILDWTSFNQERIVDVAGPGGWNDDPMLVIGNFGL--SWNQOVTQOMAL 286
DB 233 DIRRGWKAESAQNFLDWGVPTAASIIIEVR---YSIPDAFLVSGGSGIRSLDAKAIAL 288
QY 287 WAI---MAAPLFMS-----NDLRHSPOAKA---LLQDKDVAINQDP---LCK-Q 327
DB 289 GADIAGNALPVLKSAIEGKESLEQPFPRKIIFELKATMMLTGSKNVEALKRSSIVILGKLK 348
QY 328 GYQLROGDNFEVWER 342
DB 349 EWAEYRGINLSIYEK 363

RESULT 11
US-10-510-386-236
; Sequence 236, Application US/10510386
; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Olsen, Peter Bjørke
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: Improved Bacillus Host Cell
; FILE REFERENCE: 10294.204-US
; CURRENT APPLICATION NUMBER: US/10/510,386
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 236
; LENGTH: 622
```



```
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-10-510-386-236

Query Match      3.4%; Score 78.5; DB 6; Length 622;
Best Local Similarity 21.4%; Pred. No. 4.5;
Matches 65; Conservative 24; Mismatches 94; Indels 121; Gaps 16;

QY 97 APQRDSRGLQADPQRPFGIRQLANYVHSGKLGKIYADVGNKTCAGPFGSGF-----150
Db 254 ADYRNAQGRILSGPE-----YLTV-----FKGDTGABELTTV 284
QY 151 YYDIDAQTFADWG-----VDLLKFDGCYCDLEN-----LADGY-----KMSLALNRTGRSIV 199
Db 285 NYEPARGNADWDGSGYGRVDFLAGVAYLDGERPSFMARGYYTRTVLVAYNFRGCKLT 344
QY 200 YSCWPLYMPPQK--PNYTEIRQYCNHWNFNADIDDSWSKSIKSLDWTGSFNORI-----253
Db 345 -----KLWTFDSADPANGAYACQGNHSLSVADVDDGK-----DEIYIGAM 395
QY 254 -VDVAGPG-----GMNDPDMVLVGN-----FGLSMNQVVTQMALWA 288
Db 386 AVDHGKGLYSTGWGHGDAMHTGNLDPSPGLEVFQVHENSNSPYGLSPRDAKTGKIWG 445
QY 289 I-----MAAPLFMSNLRHLSPOAKA-----LLQDKQVIAINQDPLGKQGLRQGDN 336
Db 446 VHAGKQVGRGMAADI-----DPRYEGAEVWANGSLYTAKV-----KIGNTLFSSTN 492
QY 337 FEVW 340
Db 493 FGIW 496

RESULT 12
US-10-821-234-1141
; Sequence 1141, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1141
; LENGTH: 760
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1141

Query Match      3.4%; Score 78; DB 6; Length 760;
Best Local Similarity 19.2%; Pred. No. 6.6;
Matches 40; Conservative 25; Mismatches 59; Indels 84; Gaps 8;

QY 49 RFMCNLDCQEPDSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCW-----96
Db 362 RIVTYNPSAHNDANLQMDIFLEBDGNVNSKLTQDSVWNYHCWNEAMTRPDLPVGFGW 421
QY 97 -----APQRDSRGLQADP-----QRPFGIRQLANYVHSGKLGKIYADVGNKTCAGPFGS 148
Db 422 QAVDSTPQNSDGMRYRCGPASVQAIRHG-----HVCFOFDAP 458
QY 149 FGYYDIDAQTFADWGVLL-----KFDGCYC-----DSLNLADGY 184
Db 459 F-----VPAVNSDLIYITAKDQTHVENVVDATHICKLIVTKQIGDGMIDTDTY 510
QY 185 KHMS-----LALNRTGRSIVYSCWPL 206

Db 511 KFOEQGEERLAE-----TALMYGAKRPL 535

RESULT 13
US-10-467-962B-103
; Sequence 103, Application US/10467962B
; Publication No. US20050246784A1
; GENERAL INFORMATION:
; APPLICANT: Plesch, Gunnar
; APPLICANT: Blau, Astrid
; APPLICANT: Daeschner, Klaus
; APPLICANT: Klein, Mathieu
; TITLE OF INVENTION: Identification of Herbicidally Active Substances
; FILE REFERENCE: 2000 857
; CURRENT APPLICATION NUMBER: US/10/467,962B
; CURRENT FILING DATE: 2003-08-14
; PRIOR APPLICATION NUMBER: PCT/EP02/01466
; PRIOR FILING DATE: 2002-02-13
; NUMBER OF SEQ ID NOS: 109
; SOFTWARE: PatentIn Vers. 2.0
; SEQ ID NO 103
; LENGTH: 786
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-467-962B-103

Query Match      3.4%; Score 78; DB 6; Length 786;
Best Local Similarity 19.0%; Pred. No. 6.9;
Matches 76; Conservative 39; Mismatches 106; Indels 180; Gaps 20;

QY 5 NPETHLGCALALRFLALVSWDIPGARALDNGLAR-----38
Db 280 NYLTHIGHAKAM-----FVDFGLAKERGCGCYLYDDTNPEAKEEYINHI 325
QY 39 TPTMGWLHWEFMNLDQEE-----PDSCISE--K 67
Db 326 EETVKNMGWEPFKELYDLAVELIRRHAYVDHQTADDEIKREYREKKNKSPWRDRPIESLK 385
QY 68 LFMEMAELMVSEGWKADAGEYLCIDCWMAQPSRDSRGLQADPQRPFGIRQLANYVHSGK 127
Db 386 LFDMERGIIEG-----KATLRMKQDQSDNFNNYDLIAY-----421
QY 128 GLKLGKIYADVGNKTCAGPFGSGFYDIDAQTFADWGVLLKFGCYCDSLENLADGYKHM 187
Db 422 RIKFAPHPKAGDKWCI-YPS-----YD-----YAHCTVDSLENIT-----HS 457
QY 188 SLALN-RTGRSIVYSCWPL-----YMWPFQKNYTEIRQYCNHWNFNADIDDSWSKS 238
Db 458 LCTLEFETRSASY-----WLLHSLSLYMPYVMEYSRLNVT-----493
QY 239 IKSILDWTSFN-----QERIVDVAGPGWNPDMVLVGNFGLSMNQVVTQMALWAIM-----290
Db 494 -NTVMSKRKLNIYVTKYYD-----GWDDPRLTLIS--GLR-PRGYTSTAINAFVRCIGI 544
QY 291 ----AAPLFMSNLDLRI-----SPOAKALLQDKQDVIAIN 320
Db 545 TRSDGSMIHVSRLEHHIRBELNKTAPRTVMVNLPLKVITN 585

RESULT 14
US-10-131-826A-450
; Sequence 450, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
```

```

; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C128
; CURRENT APPLICATION NUMBER: US/10/131.826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 450
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-131-826A-450

```

```

RESULT 15
US-11-057-058-65
; Sequence 65, Application US/11057058
; Publication No. US2005024400A1
; GENERAL INFORMATION:
; APPLICANT: LeBowitz, Jonathan
; APPLICANT: Maga, John
; TITLE OF INVENTION: ACID ALPHA-GLUCOSIDASE AND FRAGMENTS THEREOF
; FILE REFERENCE: SYM-011
; CURRENT APPLICATION NUMBER: US/11/057,058
; CURRENT FILING DATE: 2005-02-10
; PRIOR APPLICATION NUMBER: US 60/543,812
; PRIOR FILING DATE: 2004-02-10
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 65
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Spinacia oleracea
US-11-057-058-65

```

Search completed: January 1, 2006, 00:29:52
Job time : 22.0897 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 23:15:26 ; Search time 144.752 Seconds
(without alignments)
2032.482 Million cell updates/sec

Title: US-10-602-220-16

Perfect score: 2293

Sequence: 1 MQLRNPELHLCALALRFLA.....EWTSLRSHRNPTCTVLLQL 417

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2293	100.0	429	1	P06280 homo sapien
2	2293	100.0	429	1	Q53Y83 HUMAN
3	2292	100.0	429	2	Q53HF3 homo sapien
4	2157	94.1	403	2	Q6LER7 homo sapien
5	1806	78.8	419	1	AGAL_MOUSE
6	1806	78.8	421	2	Q8GZ66 mouse
7	1480.5	64.6	429	2	Q90WL7 FUGRU
8	1267	55.3	377	2	Q4RTE7 TETNG
9	1130	49.3	338	2	Q5XJT6 BRARE
10	1074.5	46.9	431	1	Q6Q17 XENLA
11	1056.5	46.1	400	2	Q6GR44 XENLA
12	1051	45.8	405	1	NAGAB_CHICK
13	1049	45.7	412	2	Q7Q1V0 ANOGA
14	1038	45.3	410	2	Q4STX0 TETNG
15	1030	44.9	413	2	Q5VL27 DROME
16	1030	44.9	427	2	Q8MYV3 DROME
17	1007.5	43.9	419	2	Q7Q6H3 ANOGA
18	1006	43.9	417	2	Q9V7N9 DROME
19	990.5	43.2	415	1	NAGAB_MOUSE
20	982	42.8	415	1	NAGAB_RAT
21	967.5	42.2	411	1	NAGAB_BOVIN
22	952.5	41.5	411	1	NAGAB_HUMAN
23	861	37.5	451	2	Q21801_CABEL
24	849	37.0	452	2	Q60L20_CABER
25	759	32.8	399	2	Q5DBS4 SCHJA
26	729	31.8	434	2	Q9FW8 ORYSA
27	716.5	31.2	434	2	Q9LYL2 ARATH
28	716.5	31.2	437	2	Q8VXZ7 ARATH
29	712	31.1	378	1	AGAL_COFAR
30	710	31.0	466	2	Q54EG4 DICDI
31	710	31.0	478	2	Q8VW86_CLOJO

RESULT 1

ID	AGAL_HUMAN	STANDARD;	PRT;	429 AA.
AC	P06280;			
DT	01-JAN-1988	(Rel. 06, Created)		
DT	01-AUG-1988	(Rel. 08, Last sequence update)		
DT	10-MAY-2005	(Rel. 47, Last annotation update)		
DE	Alpha-galactosidase A precursor (EC 3.2.1.22) (Melibiase) (Alpha-D-			
DE	galactoside galactohydrolase) (Alpha-D-galactosidase A) (Agalsidase			
DE	alfa)			
GN	Name=GLA;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;			
OC	Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	NUCLEOTIDE SEQUENCE.			
RC	TISSUE=Lymphoblast;			
RC	MEDLINE=89263745; PubMed=2542896;			
RA	Kornreich R., Desnick R.J., Bishop D.F.;			
RA	"Nucleotide sequence of the human alpha-galactosidase A gene.";			
RL	Nucleic Acids Res. 17:3301-3302(1989).			
RN	[2]			
RP	NUCLEOTIDE SEQUENCE.			
RC	TISSUE=Fibroblast;			
RC	MEDLINE=87246603; PubMed=3036505;			
RA	Tsuji S., Martin B.M., Kaslow D.C., Migeon B.R., Choudary P.V.,			
RA	Stubblefield B.K., Mayor J.A., Murray G.J., Barranger J.A.,			
RA	Gins E.I.;			
RA	"Signal sequence and DNA-mediated expression of human lysosomal alpha-			
RT	galactosidase A.";			
RL	Eur. J. Biochem. 165:275-280(1987).			
RN	[3]			
RP	NUCLEOTIDE SEQUENCE.			
RC	MEDLINE=95352959; PubMed=7626884;			
RA	Oeltjen J.C., Liu X., Lu J., Allen R.C., Muzny D.M., Belmont J.W.,			
RA	Gibbs R.A.;			
RT	"Sixty-nine kilobases of contiguous human genomic sequence containing			
RT	the alpha-galactosidase A and Bruton's tyrosine kinase loci.";			
RL	Mamm. Genome 6:334-338(1995).			
RN	[4]			
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].			
RC	PubMed=15772651; DOI=10.1038/nature03440;			
RA	Ross M.T., Grafham D.V., Coffey A.J., Scherer S., McLay K., Muzny D.,			
RA	Platzner M., Howell G.R., Burrows C., Bird C.P., Frankish A.,			
RA	Lovell F.B., Howe K.L., Ashurst J.L., Fulton R.S., Sudbrak R., Wen G.,			
RA	Jones M.C., Hittaker A., Andrews T.D., Scott C.E., Searle S., Chen R.,			
RA	Ramsay J., Whitaker A., Deadman R., Carter N.P., Hunt S.E., Chen R.,			
RA	Cree A., Gunaratne P., Havlak P., Hodgson A., Metzker M.L.,			
RA	Richards S., Scott G., Steffen D., Sodergren E., Wheeler D.A.,			
RA	Worley K.C., Ainscough R., Ambrose K.D., Ansari-Lari M.A., Aradhy S.,			
RA	Ashwell R.I., Babbage A.K., Bagguley C.B., Ballabio A., Banerjee R.,			
RA	Barker G.E., Barlow K.F., Barrett I.P., Bates K.N., Beare D.M.,			

ALIGNMENTS

32	709	30.9	420	2	Q5DUH8_COFAR	Q5duh8 coffea arab
33	698.5	30.5	425	2	Q7XIV4_ORYSA	Q7xiv4 oryza sativ
34	697.5	30.4	431	2	Q5XTZ3_SALMI	Q5xtz3 salvia milt
35	695.5	30.3	380	2	Q9FUR8_LYCES	Q9fur8 lycopersico
36	695.5	30.3	409	2	Q9SP05_LYCES	Q9sp05 lycopersico
37	694	30.3	378	2	Q5DUH7_COFCA	Q5duh7 coffea cane
38	692	30.2	428	2	Q84VQ7_HELAN	Q84vq7 helianthus
39	687.5	30.0	417	1	AGAL_ORYSA	Q9fxt4 oryza sativ
40	687.5	30.0	422	2	Q39811_SOYBN	Q39811 glycine max
41	684.5	29.9	396	2	Q8RX86_ARATH	Q8rx86 arabidopsis
42	684	29.8	385	2	Q54AX7_DICDI	Q54ax7 dictyosteli
43	682	29.7	425	2	Q41100_PHAVU	Q41100 phaseolus v
44	681.5	29.7	411	1	AGAL_CYATE	P17479 cymatopseis t
45	681	29.7	405	2	Q5ZP79_PEA	Q5zpz79 pisum sativ

- RA Beasley H., Beasley O., Beck A., Bethel G., Blechschmidt K., Brady N.,
RA Bray-Allen S., Bridgeman A.M., Brown A.J., Brown M.J., Bonnin D.,
RA Bruford E.A., Buhay C., Burch C., Burford D., Burgess J., Burrill W.,
RA Burton J., Bye J.M., Carder C., Carrel L., Chako J., Chapman J.C.,
RA Chavez D., Chen E., Chen G., Chen Y., Chen Z., Chinault C.,
RA Ciccodicola A., Clark S.Y., Clarke G., Clee C.M., Clegg S.,
RA Clerc-Blankenburg K., Clifford K., Cobley V., Cole C.G., Conquer J.S.,
RA Corby N., Connor R.E., David R., Davies J., Davis C., Davis J.,
RA Delgado O., Deshazo D., Dhani P., Ding Y., Dinh H., Dodsworth S.,
RA Draper H., Dugand-Rocha S., Dunham A., Dunn M., Durbin K.J., Dutta I.,
RA Eades T., Ellwood M., Emery-Cohen A., Errington H., Evans K.L.,
RA Faulkner L., Francis F., Frankland J., Fraser A.E., Galgoczy P.,
RA Gilbert J., Gill R., Gloeckner G., Gregory S.G., Gribble S.,
RA Gryffiths C., Grocock R., Gu Y., Gwilliam R., Hamilton C., Hart E.A.,
RA Hawes A., Heath P.D., Heitmann K., Hennig S., Hernandez J.,
RA Hinzmann B., Ho S., Hoff M., Howden P.J., Huckle E.J., Hume J.,
RA Hunt P.J., Hunt A.R., Ischerwood J., Jacob L., Johnson D., Jones S.,
RA de Jong P.J., Joseph S.S., Keenan S., Kelly S., Kershaw J.K., Khan Z.,
RA Xioschis P., Klages S., Knights A.J., Koslira A., Kovar-Smith C.,
RA Laird G.K., Langford C., Lawlor S., Leversha M., Lewis L., Liu W.,
RA Lloyd C., Lloyd D.M., Louisedge H., Loveland J.E., Lovell J.D.,
RA Lozado R., Lu J., Lyne R., Ma J., Maheshwari M., Matthews L.H.,
RA McDowall J., McLaren S., McMurray A., Meidl P., Meitinger T.,
RA Milne S., Miner G., Mistry S.L., Morgan M., Morris S., Mueller I.,
RA Mullikin J.C., Nguyen N., Nordsiek G., Nyakatura G., O'dell C.N.,
RA Okwou G., Palmer S., Pandian R., Parker D., Parrish J.,
RA Pasternak S., Patel D., Pearce A.V., Pearson D.M., Pelan S.E.,
RA Perez L., Porter K.M., Ramsey Y., Reichwald K., Rhodes S.,
RA Ridler K.A., Schlesinger D., Schueler M.G., Sehra H.K.,
RA Shaw-Smith C., Shen H., Sheridan E.M., Shownkeen R., Skuce C.D.,
RA Smith M.L., Sotheran E.C., Steingrubber H.E., Steward C.A., Storey R.,
RA Swann R.M., Swardbeck D., Tabor P.E., Taudien S., Taylor T.,
RA Teague B., Thomas K., Thorpe A., Timms K., Tracey A., Trevanion S.,
RA Tromans A.C., d'Urso M., Verdusco D., Villaseca D., Waldron L.,
RA Wall M., Wang Q., Warren J., Warry G.L., Wei X., West A.,
RA Whitehead S.L., Whiteley M.N., Wilkinson J.E., Willey D.L.,
RA Williams G., Williams L., Williamson A., Williamson H., Wilming L.,
RA Woodmansey R.L., Wray P.W., Yen J., Zhang J., Zhou J., Zoghbi H.,
RA Zorilla S., Buck D., Reinhardt R., Pouster A., Rosenthal A.,
RA Zorilla H., Meindl A., Minx P.J., Hillier L.W., Willard H.F.,
RA Wilson R.K., Waterston R.H., Rice C.E., Vaudin M., Coulson A.,
RA Nelson D.L., Weinstein G., Suleston J.E., Durbin R., Hubbard T.,
RA Gibbs R.A., Beck S., Rogers J., Bentley D.R.;
RT "The DNA sequence of the human X chromosome.";
RL Nature 434:325-337(2005).
[5]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Uterus;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.F., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.W., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullihy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Souffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[6]
RP NUCLEOTIDE SEQUENCE OF 31-429, AND PARTIAL PROTEIN SEQUENCE.
RC TISSUE=Lung;
RX MEDLINE=86259694; PubMed=3014515;
- RA Bishop D.F., Calhoun D.H., Bernstein H.S., Hantzopoulos P., Quinn M.,
RA Deenick R.J.;
RT "Human alpha-galactosidase A: nucleotide sequence of a cDNA clone
encoding the mature enzyme.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:4859-4863 (1986).
[7]
RP NUCLEOTIDE SEQUENCE OF 1-64.
RA MEDLINE=98112869; PubMed=2892762; DOI=10.1016/0378-1119(87)90374-X;
Quinn M., Hantzopoulos P., Fidanza V., Calhoun D.H.;
RT "A genomic clone containing the promoter for the gene encoding the
human lysosomal enzyme, alpha-galactosidase A.";
RL Gene 58:177-188 (1987).
[8]
RP NUCLEOTIDE SEQUENCE OF 1-64.
RA MEDLINE=98234528; PubMed=2836863;
Bishop D.F., Kornreich R., Desnick R.J.;
RT "Structural organization of the human alpha-galactosidase A gene:
further evidence for the absence of a 3' untranslated region.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:3903-3907 (1988).
[9]
RP RNA EDITING OF POSITION 396.
RA MEDLINE=95380278; PubMed=7503918;
Novo F.J., Kruszewski A., McDermot K.D., Goldspink G., Gorecki D.C.;
RT "Editing of human alpha-galactosidase RNA resulting in a pyrimidine to
purine conversion.";
RL Nucleic Acids Res. 23:2636-2640 (1995).
[10]
RP X-RAY CRYSTALLOGRAPHY (3.45 ANGSTROMS) OF 32-422 IN COMPLEX WITH
PRODUCT, HOMODIMERIZATION, AND N-GLYCOSYLATION.
RA PubMed=15003450; DOI=10.1016/j.jmb.2004.01.035;
Garman S.C., Garbocki D.N.;
RT "The molecular defect leading to Fabry disease: structure of human
alpha-galactosidase.";
RL J. Mol. Biol. 337:319-335 (2004).
[11]
RP REVIEW ON FD VARIANTS.
RA MEDLINE=94258158; PubMed=7911050;
Eng C.M., Deenick R.J.;
RT "Molecular basis of Fabry disease: mutations and polymorphisms in the
human alpha-galactosidase A gene.";
RL Hum. Mutat. 3:103-111 (1994).
[12]
RP VARIANT FD SER-40.
RA MEDLINE=90092580; PubMed=2152885; DOI=10.1016/0014-5793(90)80046-L;
Koide T., Ishiura M., Iwai K., Inoue M., Kaneda Y., Okada Y.,
RA Uchida T.;
RT "A case of Fabry's disease in a patient with no alpha-galactosidase A
activity caused by a single amino acid substitution of Pro-40 by
Ser.";
RL FEBS Lett. 259:353-356 (1990).
[13]
RP VARIANT FD VAL-296.
RA MEDLINE=91101674; PubMed=1846223;
von Scheidt W., Eng C.M., Fitzmaurice T.F., Erdmann E., Hubner G.,
Olsen E.G.J., Christomanou H., Kandolf R., Bishop D.F., Desnick R.J.;
RT "An atypical variant of Fabry's disease with manifestations confined
to the myocardium.";
RL N. Engl. J. Med. 324:395-399 (1991).
[14]
RP VARIANT FD GLN-301.
RA MEDLINE=9102721; PubMed=2171331;
Sakuraba H., Oshima A., Fukuhara Y., Shimamoto M., Nagao Y.,
Bishop D.F., Desnick R.J., Suzuki Y.;
RT "Identification of point mutations in the alpha-galactosidase A gene
in classical and atypical hemizygotes with Fabry disease.";
RL Am. J. Hum. Genet. 47:784-789 (1990).
[15]
RP VARIANT FD TRP-356.
RA MEDLINE=89198098; PubMed=2539398;
Bernstein H.S., Bishop D.F., Astrin K.H., Kornreich R., Eng C.M.,
Sakuraba H., Desnick R.J.;
RT "Fabry disease: six gene rearrangements and an exonic point mutation
in the alpha-galactosidase gene.";

```
RL J. Clin. Invest. 83:1390-1399(1989).
RN [16]

Query Match      100.0%; Score 2293; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 7.5e-186;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MOLRNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Db 1 MOLRNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60

Qy 61 DSCISEKLFMEAEMLVSEGHKADAGEYLCIDDCWMAPOQDSERLQADPORFPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGHKADAGEYLCIDDCWMAPOQDSERLQADPORFPHGIRQL 120

Qy 121 ANYVHSGKLGIGIYADVGNKTCAGFGPSFGYYDIDAQTFADWGVDDLKFKDCYCDLSLENL 180
Db 121 ANYVHSGKLGIGIYADVGNKTCAGFGPSFGYYDIDAQTFADWGVDDLKFKDCYCDLSLENL 180

Qy 181 ADGKXMSLALNRGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKSIX 240
Db 181 ADGKXMSLALNRGRSIVYSCWPLVYMWPFQKPNYTEIROQCNHWRNFADIDDSWKSIX 240

Qy 241 SILDWTSPQKALLOQDVTAINDPGLKQGYQLRQDGFVWEPWPLSGLAWAVAMINRQIG 360
Db 241 SILDWTSPQKALLOQDVTAINDPGLKQGYQLRQDGFVWEPWPLSGLAWAVAMINRQIG 360

Qy 361 GPRSYYTIAVSLGKGVACNPACFITQLLPVKRKLGFYEWSRLRSHINPTGTVLQL 417
Db 361 GPRSYYTIAVSLGKGVACNPACFITQLLPVKRKLGFYEWSRLRSHINPTGTVLQL 417

RESULT 2
Q53Y83 HUMAN
ID Q53Y83 HUMAN PRELIMINARY; PRT; 429 AA.
AC Q53Y83
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Galactosidase, alpha.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Kainine N., Chen X., Rolfs A., Halleck A., Hines L., Eisenstein S.,
RA Koundinya M., Raphael J., Moreira D., Kelley T., LaBaer J., Lin Y.,
RA Phelan M., Farmer A.;
RT "Cloning of human full-length cDNAs in BD Creator(TM) System Donor
RT vector.";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BT006864; AAP35510.1; -; mRNA.
SQ SEQUENCE 429 AA; 48766 MW; 613F8BF21B107D7B CRC64;

Query Match      100.0%; Score 2293; DB 2; Length 429;
Best Local Similarity 100.0%; Pred. No. 7.5e-186;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MOLRNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Db 1 MOLRNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60

Qy 61 DSCISEKLFMEAEMLVSEGHKADAGEYLCIDDCWMAPOQDSERLQADPORFPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGHKADAGEYLCIDDCWMAPOQDSERLQADPORFPHGIRQL 120

Qy 121 ANYVHSGKLGIGIYADVGNKTCAGFGPSFGYYDIDAQTFADWGVDDLKFKDCYCDLSLENL 180
Db 121 ANYVHSGKLGIGIYADVGNKTCAGFGPSFGYYDIDAQTFADWGVDDLKFKDCYCDLSLENL 180
```

```
QY 181 ADGKXKSLALNRTGRSIVTSCWPLVYMWPFQKPNYTEIROYCNHWRNFADIDDSWSKSIK 240
DB 181 ADGKXKSLALNRTGRSIVTSCWPLVYMWPFQKPNYTEIROYCNHWRNFADIDDSWSKSIK 240
QY 241 SILDWTSTFQNRIRIVDVAGPGGWNDDPMLVIGNFGLSNQVQVQTMALWAIMAAPLFMSNDL 300
DB 241 SILDWTSTFQNRIRIVDVAGPGGWNDDPMLVIGNFGLSNQVQVQTMALWAIMAAPLFMSNDL 300
QY 301 RHISFOKALLQDKVDVAINQDPLGKQGYQLRQGDNFVWERPLSLGLAWAVAMINRQIEG 360
DB 301 RHISFOKALLQDKVDVAINQDPLGKQGYQLRQGDNFVWERPLSLGLAWAVAMINRQIEG 360
QY 361 GPRSYTTIYAVASLGKGVACNPACFTQLLPVKRKLGFYEWTSLRSHINPTGTVLQL 417
DB 361 GPRSYTTIYAVASLGKGVACNPACFTQLLPVKRKLGFYEWTSLRSHINPTGTVLQL 417

RESULT 4
Q6LER7_HUMAN
ID Q6LER7_HUMAN PRELIMINARY; PRT; 403 AA.
AC Q6LER7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Alpha-galactosidase A (EC 3.2.1.22) (Fragment).
GN Name=alpha-GalA;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
[1]
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Lung;
RC MEDLINE=86259694; PubMed=3014515;
RA Bishop D.P., Calhoun D.H., Bernstein H.S., Hantzopoulos P., Quinn M.,
RA Deanick R.J.;
RT "Human alpha-galactosidase A: nucleotide sequence of a cDNA clone
RT encoding the mature enzyme.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:4859-4863 (1986).
DR EMBL; D00039; BAA34059.1; -; mRNA.
DR SMR; O6LER7; 6-396.
DR GO; GO:0004557; F:alpha-galactosidase activity; IEA.
DR GO; GO:0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR002241; Glyco_hydro_27.
DR Pfam; PF02065; Melibiase; 1.
DR PRINTS; PR00740; GLHYDRASE27.
DR PRODOM; PD02572; Glyco_hydro_GHD; 1.
DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
KW Glycosidase; Hydrolase.
FT CHAIN 1..403 alpha-galactosidase A subunit peptide.
FT NON TER 1
SQ SEQUENCE 403 AA; 45804 MW; 2FE193205B8BD1A CRC64;

Query Match 94.1%; Score 2157; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 2.4e-174;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 PGARALDGLARTPTMGWLHWRPFCMLDQCEPDCISEKLFMEALMVSEGWKDAY 86
DB 1 PGARALDGLARTPTMGWLHWRPFCMLDQCEPDCISEKLFMEALMVSEGWKDAY 60
QY 87 EYLCDIDCWMAQRDSEGRQLQADQRFPHGIRQLAVYVHSGKLGIVADVGNKTCAGFP 146
DB 61 EYLCDIDCWMAQRDSEGRQLQADQRFPHGIRQLAVYVHSGKLGIVADVGNKTCAGFP 120
QY 147 GSGFYDIDAQTFADWGVLLKFGPGCVDSLENLADGKXKSLALNRTGRSIVTSCWPL 206
DB 121 GSGFYDIDAQTFADWGVLLKFGPGCVDSLENLADGKXKSLALNRTGRSIVTSCWPL 180
QY 207 YMWPFQKPNYTEIROYCNHWRNFADIDDSWSKSIKSLDWTSTFQNRIRIVDVAGPGGWNDDP 266
```

```
RESULT 5
AGAL MOUSE
ID _AGAL_MOUSE STANDARD; PRT; 419 AA.
AC P51569;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alpha-galactosidase A precursor (EC 3.2.1.22) (Melibiase) (Alpha-D-
DE galactoside galactohydrolase) (Alpha-D-galactosidase A).
GN Name=Gla; Synonyms=Ags;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Kidney;
RX MEDLINE=96125203; PubMed=8543175; DOI=10.1016/0378-1119(95)00592-7;
RA Ohshima T., Murray G.J., Nagle J.W., Quirk J.M., Kraus M.H.,
RA Barton N.W., Brady R.O., Kulkarni A.B.;
RT "Structural organization and expression of the mouse gene encoding
RT alpha-galactosidase A.";
RL Gene 166:277-280 (1995).
[2]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C129;
RX MEDLINE=95352959; PubMed=7626884;
RA Oeltjen J.C., Liu X., Lu J., Allen R.C., Muzny D.M., Belmont J.W.,
RA Gibbs R.A.;
RT "Sixty-nine kilobases of contiguous human genomic sequence containing
RT the alpha-galactosidase A and Bruton's tyrosine kinase loci.";
RL Mamm. Genome 6:334-338 (1995).
[3]
RN NUCLEOTIDE SEQUENCE.
RX MEDLINE=96316016; PubMed=8733892; DOI=10.1006/bmme.1996.0020;
RA Gotlib R.W., Bishop D.F., Wang A.M., Zeidner K.M., Ioannou Y.I.,
RA Adler D.A., DiTecthe C.M., Desnick R.J.;
RT "The entire genomic sequence and cDNA expression of mouse alpha-
RT galactosidase A.";
RL Biochem. Mol. Med. 57:139-148 (1996).
[4]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC STRAIN=FVB/N; TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.D., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Atschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
```


Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S., Gaasterland T., Gariboldi M., Giasi C., Godzik A., Gough J., Grimmer S., Gutinch S., Hirokawa N., Jackson I.J., Jarvis E.D., Kanai A., Kawaji H., Kawagawa Y., Kedzierski R.M., King B.L., Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H., Nagashima T., Numata K., Okido T., Pavan W.J., Perte G., Pesole G., Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S., Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M., Sadelin A., Schneider C., Sempile C.A., Setou M., Shimada K., Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M., Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C., Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N., Hirozane-Kikukawa T., Konno H., Nakamura M., Sakazume N., Sato K., Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S., Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A., Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J., Birney E., Hayashizaki Y.;

"Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";

Nature 420:563-573(2002).";

[4]

NUCLEOTIDE SEQUENCE.

STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;

MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;

Carninci P., Shibata Y., Hayatsu M., Sugahara Y., Shibata K., Itoh M., Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;

"Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes.";

Genome Res. 10:1617-1630(2000).

[5]

NUCLEOTIDE SEQUENCE.

STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;

MEDLINE=2050913; PubMed=11076861; DOI=10.1101/gr.152600;

Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A., Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;

"RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer.";

Genome Res. 10:1757-1771(2000).

[6]

NUCLEOTIDE SEQUENCE.

STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;

Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P., Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W., Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T., Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T., Katoh H., Kawai J., Kohji Y., Kondo S., Konno H., Kouda M., Koya S., Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M., Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y., Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H., Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tegami M., Tagawa A., Takahashi F., Takaku-Akaira S., Takeda Y., Tanaka T., Tonaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;

Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

EMBL; AK040081; BAC30508.1; -; mRNA.

DR EMBL; AK054547; BAC35819.1; -; mRNA.

DR HSSP; P06280; IR46.

DR SMR; Q8BGZ6; 34-420.

DR MGI; MGI:1347344; GlA.

DR GO; GO:0005615; C:extracellular space; TAS.

DR InterPro; IPR02241; Glyco_hydro_27.

DR InterPro; IPR00111; Glyco_hydro_GHD.

DR Pfam; PF02065; Melibiase; 1.

DR PRINTS; PR00740; GLHYDRLASE27.

DR PRODOM; PD002572; Glyco_hydro_GHD; 1.

DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.

SQ SEQUENCE 421 AA; 47844 MW; 5F66772334014B6F CRC64;

Query Match 78.8%; Score 1806; DB 2; Length 421;

Best Local Similarity 78.2%; Pred. No. 1.6e-144;

Matches 326; Conservative 41; Mismatches 50; Indels 0; Gaps 0;

QY 1 MQLNRPDLHLCALALFLALVSWDIPGARALDNGLARTPTMGLHWRFCMLDCCQEEP 60

DB 3 MKLLSRDRLVCELALCLALVFWLSILGVRLDNGLARTPTMGLHWRFCMLDCCQEEP 62

QY 61 DSCISEKLFMEALMVSQKAGYEYLICDCCWAPQDSRGLQADQRPFGHIGRL 120

DB 63 DACISEQLFMQAEMLVSDGWRDAGYDLYLCIDCCWAPERSKGRLOADQRPFGSIKHL 122

QY 121 ANYVHSGKLGIYADVGNKTCAGFGSGFYVIDAQTADGWVDLLKFGQCYCDSLENL 180

DB 123 ANYVHSGKLGIYADVGNKTCAGFGSGFSYDIDAQTADGWVDLLKFGQCHDSVSL 182

QY 181 ADGYKMSLALNRTGRSIVSYSCWPLMYPQKPNYTEIRQYCNHWNFNADIDDSWKSIIK 240

DB 183 ENGKYKMALALNRTGRSIVSYSCWPLYLRFPHKPNYTDIQYCNHWNFNDDVYDSWESIK 242

QY 241 SILDWTSTFNOERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQMALWAIWAAPLFMSNDL 300

DB 243 NILSWTVYQKEIVEVAGPGSGWNPDMLVIGNFGLSWDQVTTQMALWAIWAAPLFMSNDL 302

QY 301 RHISPOAKLLQDKDVIAINODPLGKGYQLROGDNPEVWERPLSGLAWAVAMINROBIG 360

DB 303 RQISSQNALQNDVIAINQDPLGKGYCFKRNHEVWERPLSNLAWAVAVRNLOEIG 362

QY 361 GPRSYTIAVSLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

DB 363 GPCPYTIQISSLGRGLACNPGCIITQLLPKVLHGFYEWTLTLKTRVNSPGTVLFLRL 419

RESULT 7

Q90WL7 FUGRU PRELIMINARY; PRT; 429 AA.

AC Q90WL7;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Alpha galactosidase a.

GN Name=agal;

OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).

OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percomorphia; Tetraodontiformes;

OC Tetraodontidae; Tetraodontidae; Takifugu.

OX NCBI_TaxID=31033;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RA Goode D., Elgar G.;

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

[2]

RP NUCLEOTIDE SEQUENCE.

RA Elgar G.S.;

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ290422; CAC4626.1; -; Genomic DNA.

DR HSSP; P06280; IR46.

DR Ensembl; SINFRUG00000147523; Fugu rubripes.

DR GO; GO:0004553; P:hydrolyase activity, hydrolyzing O-glycosyl . . . ; IEA.

DR GO; GO:0005975; P:carbohydrate metabolism; IEA.

DR InterPro; IPR002241; Glyco_hydro_27.

DR InterPro; IPR000111; Glyco_hydro_GHD.

DR Pfam; PF02065; Melibiase; 1.

DR PRINTS; PR00740; GLHYDRLASE27.

DR PRODOM; PD002572; Glyco_hydro_GHD; 1.

DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.

SQ SEQUENCE 429 AA; 48480 MW; 536A826D520CF01C CRC64;

Query Match 64.6%; Score 1480.5; DB 2; Length 429;

Best Local Similarity 68.4%; Pred. No. 6.7e-117;

Matches 271; Conservative 50; Mismatches 72; Indels 3; Gaps 2;
 QY 20 ALVSWDIPGAPALDNGLARTPTMGWLHWRFCNLDCCQEPDSCISEKLFMEAEMLVSE 79
 DB 7 ALVFLIGPAAEALDNGLALPTMGWLHWRFCNLDCCQEPDSCISEKLFMEAEMLVSE 66
 QY 80 GWKDAAGYELCIDCWMAPOQDSRGLQADPQPHGIRQLANVHSGKLGKGIYADGVN 139
 DB 67 GWKDAAGYELCIDCWMAPOQDSRGLQADPQPHGIRQLANVHSGKLGKGIYADGVN 126
 QY 140 KTCAGFGSGFYDIDDAQTFADGWGVDLLKFGDCYCDLSLENLADGKYMMSLALNRTGRSIV 199
 DB 127 KTCAGFGSGFYDIDDAQTFADGWGVDLLKFGDCYCDLSLENLADGKYMMSLALNRTGRSIV 185
 QY 200 YSCWPLMYMPFQKPNYTEIRQYCNHWRNFADIDDSWKSISILDTWTSFNOERIVDVAGP 259
 DB 186 YSCWPLMYMPFQKPNYTEIRQYCNHWRNFADIDDSWKSISILDTWTSFNOERIVDVAGP 245
 QY 260 GGWNPDMVLVGNFGLSNQVQVOTOMALWMAAFLFMSNDRHISPOAKALLQDKVIAI 319
 DB 246 GGWNPDMVLVGNFGLSNQVQVOTOMALWMAAFLFMSNDRHISPOAKALLQDKVIAI 305
 QY 320 NQDPLGKQGYQLRGDNFVWERPLSLGLAWAVAMINRQEIIGRPSYTIIVASLKGKGVACN 379
 DB 306 NQDPLGKQGYQLRGDNFVWERPLSLGLAWAVAMINRQEIIGRPSYTIIVASLKGKGVACN 363
 QY 380 PACFITQLPVKRLGKGYEWTSLRSHINPTGTVLL 415
 DB 364 PACNVTQILPHYKDMQVPMFVKMVRVNPSTVLL 399

RESULT 8

Q4RTE7 TETNG
 ID Q4RTE7 TETNG PRELIMINARY; PRT; 377 AA.
 AC Q4RTE7;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Chromosome 1 SCAP14998, whole genome shotgun sequence.
 GN ORFNames=gstng0029285001;
 OS Tetraodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 CX NCBI_TaxID=99883;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Doseat C., Segurens B.,
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
 RA Bimont C., Skalli Z., Catolico L., Poulain J., De Berardinis V.,
 RA Cruaud C., Duprat S., Brottier P., Coutancan J.P., Gouzy J.,
 RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
 RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
 RA Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.;
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 RT the early vertebrate proto-karyotype.";
 RL Nature 431:946-957(2004).
 RN [2]

NUCLEOTIDE SEQUENCE

RG Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; CAE01014998; CAC08335.1; -; Genomic DNA.
 SQ SEQUENCE 377 AA; 42048 MW; 236F4E5A95B9F1A8 CRC64;

Query Match 55.3%; Score 1267; DB 2; Length 377;
 Best Local Similarity 72.1%; Pred. No. 7.5e-99;
 Matches 227; Conservative 38; Mismatches 48; Indels 2; Gaps 2;

QY 16 LRFALALSWDI-PCARALDNGLARTPTMGWLHWRFCNLDCCQEPDSCISEKLFMEAE 74
 DB 3 LYLFIILVLFVSGPAVALDNGLALPTMGWLHWRFCNLDCCQEPDSCISEKLFMEAE 62
 QY 75 LMSVSEGWKDAAGYELCIDCWMAPOQDSRGLQADPQPHGIRQLANVHSGKLGKGIY 134
 DB 63 VMVKEGWKEAGYEVVICDDCWPSRRNIFGRQLQADPKRFPGGIKKLAHYVHSRGLKGIY 122
 QY 135 ADVGNKTCAGFPSPGYVDIDAQTFADGWGVDLLKFGDCYCDLSLENLADGKYMMSLALNRT 194
 DB 123 ADVGSLTCAAGYPSGLGYDIDAQTFADGWGVDLLKFGDCYCDLSLENLADGKYMMSLALNRT 181
 QY 195 GRSTVYSCWPLMYMPFQKPNYTEIRQYCNHWRNFADIDDSWKSISILDTWTSFNOERIV 254
 DB 182 GRSTVYSCWPLMYMPFQKPNYTEIRQYCNHWRNFADIDDSWKSISILDTWTSFNOERIV 241
 QY 255 DVAGPGGWNPDMLVIGNFGLSNQVQVOTOMALWMAAFLFMSNDRHISPOAKALLQDK 314
 DB 242 PSAGPGGWNPDMLVIGNFGLSHDQESQWALWMAAFLFMSNDRHISPOAKALLQDK 301
 QY 315 DVTAINDQPLGKQGY 329
 DB 302 HVTAINDQPLGKQGF 316

RESULT 9

Q5XJT6 BRARE
 ID Q5XJT6 BRARE PRELIMINARY; PRT; 338 AA.
 AC Q5XJT6;
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Zgc:101584.
 GN ORFNames=zgc:101584;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 CX NCBI_TaxID=7955;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RA Director MGC Project;
 RL Submitted (OCT-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC083209; AAH83209.1; -; mRNA.

DR ZFIN; ZDB-GENE-041010-207; zgc:101584.
 DR GO: 0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
 DR GO: 0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR002241; Glyco_hydro_27.
 DR InterPro; IPR000111; Glyco_hydro_GHD.
 DR PRINTS; PR00740; GLHYDRLASE27.
 DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
 SQ SEQUENCE 338 AA; 37846 MW; A3FC5C461B7C9649 CRC64;

Query Match 49.3%; Score 1130; DB 2; Length 338;
 Best Local Similarity 60.3%; Pred. No. 2.8e-87;
 Matches 210; Conservative 47; Mismatches 75; Indels 16; Gaps 2;

QY 68 LFEMAEALMWSEGWKADAGEYLICDDCWMAQPDSEGRQLQADPQRPFGIRQLANYVHSK 127
 DB 2 LMSLCALMIA-GLRNKG-----THRGACRQTPKGFPPVASKKLLADYVHSK 45

QY 128 GLKLGIVADVGNKTCAGFPGSFGYDIDAQTFADWGVLLKFDGCYCDLSLENLADGKHM 187
 DB 46 GLKLGIVADVGTCTCAGPGLSGLYYDIDAKTFADWGVLLKFDGCFMPDWHQLGEGYINM 105

QY 188 SLALNRTGRSIVSCWPLMYWPKNYTEIROYCNHWRNFADIDDSWKSIIKLDWTS 247
 DB 106 SSALNQTGRSIVSCWPLMYWPKNYTEIROYCNHWRNFADIDDSWKSIIKLDWTS 165

QY 248 FNQERIVDVAGPGGWNDDMLVIGNFGLSNQVQTMALWALMAAPLFMSNDLRLHISPOA 307
 DB 166 EKQIVVPVAGPGGWNDDMLVIGNFGLSNQVQTMALWALMAAPLFMSNDLRLHISPOA 225

QY 308 KALLQDKVDIAINQDPLGKQYQLRQGNFVWRPLSGLAWAMINROEIGSPRSYTI 367
 DB 226 KELLQNKQIIAINDQDPLGKQYRLKADSFEWPLSGNRLAVAVNQRQEIIGSPRFTI 285

QY 368 AVASLGKGVACNPACFTQLLPVVRKLGFEVWTSRLRSHINPTGTVL 415
 DB 286 SVAIMPWSKLCNPKCNVTQILPTYKENGQVLLSEVVVVQVNPQTGTILL 333

RESULT 10
 Q6GQ17_XENLA PRELIMINARY; PRT; 431 AA.
 AC Q6GQ17;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE LOC443592 protein (fragment).
 GN Name=LOC443592;
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
 OC Xenopodinae; Xenopus; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RX MEDLINE=22348257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzyzanski M.I., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
 RA Richardson P.,
 RA "Genetic and genomic tools for Xenopus research: The NIH Xenopus
 RT initiative.";
 RL Dev. Dyn. 225:384-391(2002).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Embryo;
 RA Klein S., Gerhard D.S.; to the EMBL/GenBank/DBJ databases.
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC072931; AAH72931.1; -; mRNA.
 DR GO: 0004553; P:hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
 DR GO: 0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR002241; Glyco_hydro_27.
 DR InterPro; IPR000111; Glyco_hydro_GHD.
 DR Pfam; PF02065; Melibiase; 1.
 DR PRINTS; PR00740; GLHYDRLASE27.
 DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
 DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
 FT NON TER 1
 SQ SEQUENCE 431 AA; 48787 MW; 4C578DB433983471 CRC64;

Query Match 46.9%; Score 1074.5; DB 2; Length 431;
 Best Local Similarity 49.3%; Pred. No. 1.9e-82;
 Matches 209; Conservative 62; Mismatches 126; Indels 27; Gaps 6;

QY 4 RNFELHLGCALARFLALVSWDIPGARALDGLARTPTMGLHWERFMCNLDCCOEPDSC 63
 DB 18 RGYRMLHTSLCLLTLMGLCW-----CLDNLGLVTPPMGMMTWQTRCNDICKSDPDNC 71

QY 64 ISEKLFMEAEALMWSEGWKADAGEYLICDDCWMAQPDSEGRQLQADPQRPFGIRQLANY 123
 DB 72 ISENLIKIMADEMADSCWRDLGVYVICDDCWQKQSDNSGRQLQDPFERFSGMKALADY 131

QY 124 VHSKGLKLGIVADVGNKTCAGFPG-SFGYYDIDAQTFADWGVLLKFDGCYCDLSLENLAD 182
 DB 132 VHAKGLKLGIVSDMGNYTCGGYPTGLDTIKTDAETPASWEVDMLKFDGCYSNSTER-AL 190

QY 183 GYKXMSLALNRTGRSIVSCWPLMYWPKF-QKNYTEIROYCNHWRNFADIDDSWKSIIKS 241
 DB 191 GYPMSEALNATGRPILYSCWPAFEGGLPPKVNVTOLGSCNMWRNYDDIDDSWSDVLD 250

QY 242 ILDWTSTNQERIVDVAGPGGWNDDMLVIGNFGLSNQVQTMALWALMAAPLFMSNDLR 301
 DB 251 IIEWAKNQDVLQPAAGFGHWNDDMLITGDFGLSYEQSKSQAIAWAILAAPLMSNDLR 310

QY 302 HISPOAKALLQDKVDIAINQDPLGKQYQLRQGNFVWRPLSGLAWAMINROEIGG 361
 DB 311 AISQDAKDLLQNLRIIYNQDSLKQKQSLISRGISNLEWVKRELINGQYALAVLNKGTDLG 370

QY 362 PRSVTTAVASLGKGVACNPACFTQLLP-----VKRKLGFVWTSRLRSHINPTGTV 413
 DB 371 PRPYSTNLGLLN-----VTQTDGYKLVNVEFEYLMGFTSTPTPIIRVNPTGVI 420

QY 414 LLQL 417
 DB 421 FLFL 424

RESULT 11
 Q6GR44_XENLA PRELIMINARY; PRT; 400 AA.
 ID Q6GR44;
 AC Q6GR44; 2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)

```

DE GN MGC81044 protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Ovary;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Matra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Ovary;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391 (2002).
[3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Ovary;
RA Klein S., Gerhard D.S.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC071089; AAH71089.1; -; mRNA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR002241; Glyco_hydro_27.
DR Pfam; PF02065; Melibiase; I.
DR PRINTS; PR00740; GLHYDRASE27.
DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
SQ SEQUENCE 400 AA; 44854 MW; CBC4770F613FF3BA CRC64;

Query Match 46.1%; Score 1056.5; DB 2; Length 400;
Best Local Similarity 51.0%; Pred. No. 5.8e-81;
Matches 201; Conservative 58; Mismatches 114; Indels 21; Gaps 5;

Qy 32 LDNGARTPTTGMWHLWRFMNCNLCQBEPSDCISEKLFMEAEMLVSEGKWDAGYEYLCI 91
Db 7 LDNGLVTPPQMTWQRYRCNICKGDPDNCISENLKSMADKADSGWRDVGTYTISI 66

Qy 92 DDCWMAQRDSEGLQADPQPFHGIQRLANYVHSKGLGIYADVGNKTCAGPFG-SFG 150
Db 67 DDCWSQQRDSNGRLQDPLERFPFGMKALADYVHVAKGLGIYDWMGTTCGYPGGTLD 126

Qy 151 YYDIDAOTFADGWVDLLKFDQCYCDLSLENLADGKHKMSLALNRTGRSVSCSEWPLYMWP 210
Db 127 TIKIDATFASWEVDMLKFDQCYCNSNTEK-ALGPKMSSEALNGTRFILXSCSWPAYEGG 185

Qy 211 F-QKPNYTEIRQYCNWRNFADIDDSWKSISILDWTSFNQERIVDVAGPGGWNDDPMLV 269

```

```

Db 186 LPPKVNVTQSGSICNMWRNTYGDQSDWSVDLIIWYAKQODVLQPAAGPGRWNPDKLI 245
Qy 270 IGNFGLSWNQVQTOMALWAIWMAAPLFMSNDLRHISPOKALLQDKDVIAINQDPLGQGY 329
Db 246 TGFGLSYEQSKSLAIWAILAALPLINSNDLRTISQADKOLLQNRLLIYNQDALGKQGS 305
Qy 330 QLRQGDNFVWERPLSLGAWAVAMINRQETGGPRSYTIVASLGKGVACNPACFITQLLP 389
Db 306 LISQVGSLEVYKRELINGQYAVAVLNKGTDLGPRPTTSLGLN-----ITQCTD 355
Qy 390 -----VRKGLGYEWTSLRSHINPTGTVLL 415
Db 356 GYKMYNVFEKYEYIGMFKSGTIDMRVNPVTGVIFL 389

RESULT 12
NAGAB_CHICK
ID NAGAB_CHICK STANDARD; PRT; 405 AA.
AC Q90744;
DT 13-SEP-2005 (Rel. 48, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Alpha-N-acetylgalactosaminidase (EC 3.2.1.49) (Alpha-galactosidase B).
GN Name=NAGA;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], AND PARTIAL PROTEIN SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=94060104; PubMed=8241271; DOI=10.1016/0167-4781(93)90158-A;
RA Davis M.O., Hata J., Smith D., Walker J.C.;
RT "Cloning and sequence of a chicken alpha-N-acetylgalactosaminidase
RT gene.";
RL Biochim. Biophys. Acta 1216:296-298 (1993).
[2]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX PubMed=12005440; DOI=10.1016/S0969-2126(02)00726-8;
RA Garman S.C., Hannick L., Zhu A., Garbocki D.N.;
RT "The 1.9 A structure of alpha-N-acetylgalactosaminidase: molecular
RT basis of glycosidase deficiency diseases.";
RL Structure 10:425-434 (2002).
CC -| CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing N-acetyl-
CC D-galactosamine residues in N-acetyl-alpha-D-galactosaminides.
CC -| SUBCELLULAR LOCATION: Lysosomal (By similarity)
CC -| SIMILARITY: Belongs to the glycosyl hydrolase 27 family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; L18754; AAA16614.1; -; mRNA.
DR PIR; S45522; S45522.
DR PDB; 1KTB; X-ray; A=1-405.
DR PDB; 1KTC; X-ray; A=1-405.
DR Ensemble; ENSGALG0000011900; Gallus gallus.
DR InterPro; IPR002241; Glyco_hydro_27.
DR InterPro; IPR000111; Glyco_hydro_GHD.
DR Pfam; PF02065; Melibiase; I.
DR PRINTS; PR00740; GLHYDRASE27.
DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
KW 3D-structure; direct protein sequencing; Glycoprotein; Glycosidase;
KW Hydrolase; Lysosome.
FT CHAIN 1 405 Alpha-N-acetylgalactosaminidase.
FT ACT_SITE 140 140 Nucleophile (By similarity).
FT ACT_SITE 201 201 Proton donor (By similarity).
FT CARBOHYD 161 161 N-linked (GlcNAc...).
```

```
FT CARBOHYD 185 185 N-linked (GlcNAc. . .).
FT CARBOHYD 369 369 N-linked (GlcNAc. . .).
FT DISULFID 21 63
FT DISULFID 25 32
FT DISULFID 111 142
FT DISULFID 171 193
SQ SEQUENCE 405 AA; 45615 MW; E1EC0061739C305C CRC64;

Query Match 45.8%; Score 1051; DB 1; Length 405;
Best Local Similarity 53.0%; Pred. No. 1.7e-80;
Matches 206; Conservative 58; Mismatches 117; Indels 8; Gaps 6;

Qy 32 LDNGLARTPTMGWLHWRFCMCLDCEEPDSCISEKLFMEAEMLVSEGWKDAGYEYLCI 91
Db 1 LENGARTPPMGLAWERFCNVNCRDPDPCISSEMLFMEADRIADGRELGYKINI 60

Qy 92 DDCWMAQRDSRGLQADPQRPFGIRQLANYVHSGKLGKIYADVGNKTCAGPPG-SFG 150
Db 61 DDCWAAQRDAEGRVLPDPERFPRGIRKALADYVHARGLKLDIYGLGRLTCGGYPGTTLD 120

Qy 151 YYDIDAQTFADWGDLLKPDGCGYCDLSLENLADGKHSALNRTGRSIVVSCWPLYMWP 210
Db 121 RVEQDAQTFADWGDLLKPDGCGYCDLSLENLADGKHSALNRTGRSIVVSCWPLYMWP 210

Qy 211 F-QRPNTYETROYCNHWRNFADIDDSWKSISILDWTSFNQERIVDVAGPGWNDPMLV 269
Db 180 LPPKNTYTLGEICNLWRNYDDIQQSDSVLSIVDWFNTQDVLPQFAGPGHWNDDPMLI 239

Qy 270 IGNGLSWNOQVOTOMALWAIMAAPLPMNDLRHISPOAKALLQDKOVIAINQDPLGKGY 329
Db 240 IGNGFLSYEQSRSQMALWTMAAPLLMSTDLRTISPSAKTILQNLMIQINQDPLGIQGR 299

Qy 330 Q-LRQGNFVFWERPLSLGLAWAVAMINRQIEGGPRSYTIAVASLGKGVACNPACFITOLL 388
Db 300 RIIEGSHIEVFLRPLSQASALVFFSRR-TDMPFRYTTSLAKLGPFMG---AAVEYQDV 355

Qy 389 PVKRLGFEWTSRLRSHINPTGVLLQL 417
Db 356 YSGKIISGLKGDNTFVIINPFGVVMWYL 384

RESULT 13
Q01V0 ANOGA PRELIMINARY; PRT; 412 AA.
AC Q01V0;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ENSANGP0000020847 (Fragment).
GN ORFNames=ENSANGG0000018358;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Culicidae;
OC Anophelinae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG "Anopheles gambiae Sequence Committee;
RT "Anopheles gambiae re-annotation.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008980; EAL14548.2; -; Genomic DNA.
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR002241; Glyco_hydro_27.

DR InterPro; IPR000111; Glyco_hydro_GHD.
DR Pfam; PF02065; Melibiase_I.
DR PRINTS; PR00740; GLHYDRLASE27.
DR ProDom; PD002572; Glyco_hydro_GHD; 1.
FT NON TER 1
FT NON TER 412
SQ SEQUENCE 412 AA; 46517 MW; 4C91C8FB9EA6C88F CRC64;

Query Match 45.7%; Score 1049; DB 2; Length 412;
Best Local Similarity 49.2%; Pred. No. 2.6e-80;
Matches 193; Conservative 71; Mismatches 118; Indels 10; Gaps 4;

Qy 31 ALDNGLARTPTMGWLHWRFCMCLDCEEPDSCISEKLFMEAEMLVSEGWKDAGYEYLCI 90
Db 17 SLENGLARTPPMGLSWERFCNTDCEGDPENCISEHLFTTMADLVVSEGYAAVGYEYIN 76

Qy 91 IDCWMAQRDSRGLQADPQRPFGIRQLANYVHSGKLGKIYADVGNKTCAGPPGSGFG 150
Db 77 VDCWLEKSRGPRGELVADRRRFPFGMKALANYVHAKGLKFGIYEDYGNKTCAGYPGILG 136

Qy 151 YYDIDAQTFADWGDLLKPDGCGYCDLSLENLADGKHSALNRTGRSIVVSCWPLY-MW 209
Db 137 FSANDAQAQFASWDVYVYKLDGCSLPID-MDHGYPEPGRNLNATGRPMIYSCSNPVYQIY 195

Qy 210 PFOKPNYETROYCNHWRNFADIDDSWKSISILDWTSFNQERIVDVAGPGWNDPMLV 269
Db 196 AGNPNYSSIIQHNLWRNYDDIQQSDSVLSIIDYGNNDALIPNAGPGHWNDDPMLI 255

Qy 270 IGNGLSWNOQVOTOMALWAIMAAPLPMNDLRHISPOAKALLQDKOVIAINQDPLGKGY 329
Db 256 IGNGFLSYEQSKTOMALWAIMAAPLPMNSVDLRTIRPEFKAILQNRKIIAVDQDPLGTQGR 315

Qy 330 QLRQGNFVFWERPLSLGL---AWAVAMINRQIEGGPRSYTIAVASLGKGVACNPACFI 384
Db 316 RIYKHGIEIWSRPITPIYQTYTYSYAFNRRRTDGTFSVAVTLRELG---LIISPTGYR 372

Qy 385 TOLLPVKRLGFEWTSRLRSHINPTGVLLQL 416
Db 373 VEDLYEVDYGVLSPTQTKIKVKNPVGVLIR 404

RESULT 14
Q4STX0 TETNG PRELIMINARY; PRT; 410 AA.
AC Q4STX0;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Chromosome undetermined SCAFI4084, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG00012743001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anhouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Blemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Parra G., Duprat S., Brottier P., Coutanceau J.P., Guzy J.,
RA Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Winkler P., Lander E.S., Weissbach J., Roest Crolius H.;
RA "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
```

```

RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAAB01014084; CAP9512.1; -; Genomic_DNA.
FT NON TFR 410
SQ SEQUENCE 410 AA; 46172 MW; F5C661DA72DCEB88 CRC64;

Query Match 45.3%; Score 1038; DB 2; Length 410;
Best Local Similarity 48.9%; Pred. No. 2.2e-79;
Matches 207; Conservative 58; Mismatches 116; Indels 42; Gaps 8;

QY 8 LHLCALALFLALVSWDIPGARLNDGLARTPMGLHWRFCNLDCEEPDSCISEK 67
Db 1 MHLALLLASVLSL-----GTLALDNLGLWRTPPMGWLAWERFACDIDCDQPFNCISEN 54
QY 68 LFMEAMELVSEGKADAGYBYLCIDDCWMAQPDSEGRLOADQRFPHGIRQLANYVHSK 127
Db 55 LFTDMADRLQDQGWKDLGYVYVNIIDCWSSKERDKKGRLOPDKRPPGGIPKLARYMHR 114
QY 128 GLKGIYADVGNKTCAGFPQS-FGYDIDAQTPADWGVLLKFDGCVDSLENLADGYKH 186
Db 115 GLKGIYDGMGTLTCGGYPTGLDKIDIDAKTFAEWEDVLMKFDGCVSNEVQ-QQGYPL 173
QY 187 MSALANTGRSIVYSCWPLMYWPF-QKPNYETIRQYCNHWRNPAFDDSKWKSITLDW 245
Db 174 MSXALNATGRPIAYSCSWPAYSGGLPPKVNVTQGEICNLWRNYDDIQDSWDSVNLIDW 233
QY 246 TSFNQERIVDVGPGWNDPDMVLVIGNFGLSWNQVQTMALMAIMAPLWMSNDRHISP 305
Db 234 FFDNQDIVAPAGGRWNDPDMVLVIGNFGLSQRSTQMALMAIMAPLWMSNDRITISA 293
QY 306 QAKALQKQDVIAINQPLQKQVQL-RQGNFVFWERPLSGLAAMAVAMINROEIGGPRS 364
Db 294 EARSILQKIAIGINQPLQFGQRRLVKESGIEVFWRLSNDTSALVFFSRR----- 346
QY 365 YTTAVASLHGKVCNPACTTOLLVPRKLGFEY-----WTSRLRS-----HINPTG 411
Db 347 -----TDPYRYKTSLSKUSYPTGYKIVDVFTQKGSLLKDSADFVVSVPNTG 394

QY 412 TVL 414
Db 395 VVM 397

RESULT 15
Q9VL27 DROME
ID Q9VL27 DROME PRELIMINARY; PRT; 413 AA.
AC Q9VL27;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CG5731-PA.
GN Names=CG5731; ORFNames=CG5731;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

```

```

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Belcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kemnison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirkas R., Tabor P.B., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RT melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirkas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective."
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RG Berkeley Drosophila Genome Project;
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirkas R., Smith E.,
RA Yu C., Rubin G.;
RT "Drosophila melanogaster release 4 sequence."
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 21:05:01 ; Search time 152.421 Seconds
(without alignments)
1202.074 Million cell updates/sec

Title: US-10-602-220-16
Perfect score: 2293
Sequence: 1 MQRNPELHGCAALRFLA.....EWTSLRSHINPTGTVLLQL 417

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_21.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2293	100.0	417	5	Aae28212 Human rga
2	2293	100.0	417	7	Add84755 Human alp
3	2293	100.0	417	8	Adj88281 Human wt
4	2293	100.0	417	8	Adm48685 Human wil
5	2293	100.0	417	8	Adu65920 Human alp
6	2293	100.0	417	9	Aea27449 Human alp
7	2293	100.0	421	5	Aae28210 Human rga
8	2293	100.0	421	7	Add84751 Human alp
9	2293	100.0	421	8	Adj88277 Human wt
10	2293	100.0	421	8	Adm48681 Human wil
11	2293	100.0	421	8	Adu65916 Human alp
12	2293	100.0	421	9	Aea27445 Human alp
13	2293	100.0	423	5	Aae28213 Human rga
14	2293	100.0	423	7	Add84757 Human alp
15	2293	100.0	423	8	Adj88283 Human wt
16	2293	100.0	423	8	Adm48687 Human wil
17	2293	100.0	423	8	Adu65922 Human alp
18	2293	100.0	423	9	Aea27451 Human alp
19	2293	100.0	425	9	Aea27472 Human alp
20	2293	100.0	427	5	Aae28211 Human rga
21	2293	100.0	427	7	Add84753 Human alp
22	2293	100.0	427	8	Adj88279 Human wt
23	2293	100.0	427	8	Adm48683 Human wil
24	2293	100.0	427	8	Adu65918 Human alp

25	2293	100.0	427	9	Aea27447 Human alp
26	2293	100.0	429	2	Aar53765 Alpha-gal
27	2293	100.0	429	2	Aar63234 Human alp
28	2293	100.0	429	2	Aar70207 Alpha-gal
29	2293	100.0	429	5	Aae28206 Human wt
30	2293	100.0	429	7	Adf76604 Novel hum
31	2293	100.0	429	8	Adn49740 Human alp
32	2293	100.0	429	8	Adp24703 PRO poly
33	2293	100.0	429	8	Adu18066 Human can
34	2293	100.0	429	8	Adu74416 Human alp
35	2293	100.0	429	9	Ady19459 PRO poly
36	2293	100.0	429	9	Ady17095 PRO poly
37	2293	100.0	429	9	Aea27471 Human alp
38	2293	100.0	429	9	Aeb43188 Human alp
39	2293	100.0	430	7	Abm79001 Human alp
40	2293	100.0	431	5	Aae28209 Human rga
41	2293	100.0	431	7	Add84749 Human alp
42	2293	100.0	431	8	Adj88275 Human wt
43	2293	100.0	431	8	Adm48679 Human wil
44	2293	100.0	431	8	Adu66914 Human alp
45	2293	100.0	431	9	Aea27443 Human alp

ALIGNMENTS

RESULT 1
AAE28212
ID AAE28212 standard; protein; 417 AA.
XX
AC AAE28212;
XX
DT 27-DEC-2002 (first entry)
XX
DE Human rGAL-12 protein.
XX
KW Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
KW therapeutic; rGAL-12.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Cleavage-site 401..402 /note= "CTPP cleavage site"
FT
XX
PN US2002088024-A1.
XX
PD 04-JUL-2002.
XX
PF 13-NOV-2001; 2001US-00993059.
XX
PR 26-JUL-2000; 2000US-00626127.
XX
PA (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
XX
PI (KUMA/) KUMAGAI M H.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2002-681656/73.
XX N-PSDB; AAD45225.
XX
XX Novel human alpha-galactosidase polypeptide useful for treating lysosomal storage diseases.
XX
XX Claim 7; Page 49-50; 88pp; English.
XX
XX The invention relates to human alpha-galactosidase truncated at the carboxy terminus and the production of enzymatically active recombinant human and animal lysosomal enzymes. The invention is useful for producing lysosomal enzymes for treating lysosomal storage diseases, producing altered or mutated proteins, enzymatically active or otherwise, to serve as precursors or substrates for further in vivo or in vitro processing to

CC a specialised industrial form for research or therapeutic uses, to
 CC produce more effective therapeutic enzyme, for producing antibodies
 CC against lysosomal enzymes for medical diagnostic use, and in any
 CC commercial process that involves substrate hydrolysis. The present
 CC sequence is human rGAL-12 protein
 XX
 SQ Sequence 417 AA;

Query Match 100.0%; Score 2293; DB 5; Length 417;
 Best Local Similarity 100.0%; Pred. No. 5.3e-220;
 Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MOLRNPELHIGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFRCMNCDCQEEP 60
 DB 1 MOLRNPELHIGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFRCMNCDCQEEP 60
 QY 61 DSCISEKLFMEAEMLVSEGWKDGAYEYLICDDCWMAQPDSEGRLOADPQRPFGHGIQRL 120
 DB 61 DSCISEKLFMEAEMLVSEGWKDGAYEYLICDDCWMAQPDSEGRLOADPQRPFGHGIQRL 120
 QY 121 ANYVHSGKLGKIYADVGNKTCAGFPGSGYGYDIDAQTFADWGVDDLKFPDGCYCDLSLENL 180
 DB 121 ANYVHSGKLGKIYADVGNKTCAGFPGSGYGYDIDAQTFADWGVDDLKFPDGCYCDLSLENL 180
 QY 181 ADGYKMSLALNRTGRSIVVSCWPLWMPFQKPNTEIRQYCNHWRNFADIDDSWKSIIK 240
 DB 181 ADGYKMSLALNRTGRSIVVSCWPLWMPFQKPNTEIRQYCNHWRNFADIDDSWKSIIK 240
 QY 241 SILDWTFSNQERIVDVAGPGGNDPDMLVIGNFGLSNQOVTQMALWAINAAPLFMSNDL 300
 DB 241 SILDWTFSNQERIVDVAGPGGNDPDMLVIGNFGLSNQOVTQMALWAINAAPLFMSNDL 300
 QY 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRGDNFEVWERPLSLGLAWAVAMINRQEI 360
 DB 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRGDNFEVWERPLSLGLAWAVAMINRQEI 360
 QY 361 GPRSYTIAVASLGKGVACNACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
 DB 361 GPRSYTIAVASLGKGVACNACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 2
 ADD84755
 ID ADD84755 standard; protein; 417 AA.

XX
 AC ADD84755;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Human alpha-galactosidase rGAL-12 polypeptide.
 XX
 KW Human; alpha-galactosidase; rGAL-4; lysosomal enzyme;
 KW enzyme replacement therapy; lysosomal disease; enzyme.
 XX
 OS Homo sapiens.
 XX
 PN US2003106095-A1.
 XX
 XX 05-JUN-2003.
 XX
 XX 20-MAR-2002; 2002US-00103327.
 XX
 XX 26-JUL-2000; 2000US-00626127.
 PR 13-NOV-2001; 2001US-00993059.
 XX
 XX (GARG/) GARGER S J.
 PA (TURP/) TURPEN T H.
 PA (KOMA/) KOMAGAI M H.
 XX
 XX Garger SJ, Turpen TH, Kumagai MH;
 PI
 XX WPI; 2003-801257/75.
 DR N-PSDB; ADD84754.
 DR

XX New polynucleotide for producing active recombinant human and animal
 PT lysosomal enzymes in a plant expression system that can be used in enzyme
 PT replacement therapy.
 XX
 PS Claim 7; SEQ ID NO 16; 77pp; English.
 XX

CC The invention relates to human alpha-galactosidase derivatives and the
 CC nucleic acids encoding them. The polypeptides are used in a method for
 CC producing active recombinant human and animal lysosomal enzymes in a
 CC plant expression system. The enzymes can be used in enzyme replacement
 CC therapy for the therapeutic treatment of human and animal lysosomal
 CC diseases. This sequence represents a human alpha-galactosidase derivative
 CC polypeptide of the invention.
 XX

SQ Sequence 417 AA;

Query Match 100.0%; Score 2293; DB 7; Length 417;
 Best Local Similarity 100.0%; Pred. No. 5.3e-220;
 Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLRNPELHIGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFRCMNCDCQEEP 60
 DB 1 MOLRNPELHIGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFRCMNCDCQEEP 60
 QY 61 DSCISEKLFMEAEMLVSEGWKDGAYEYLICDDCWMAQPDSEGRLOADPQRPFGHGIQRL 120
 DB 61 DSCISEKLFMEAEMLVSEGWKDGAYEYLICDDCWMAQPDSEGRLOADPQRPFGHGIQRL 120
 QY 121 ANYVHSGKLGKIYADVGNKTCAGFPGSGYGYDIDAQTFADWGVDDLKFPDGCYCDLSLENL 180
 DB 121 ANYVHSGKLGKIYADVGNKTCAGFPGSGYGYDIDAQTFADWGVDDLKFPDGCYCDLSLENL 180
 QY 181 ADGYKMSLALNRTGRSIVVSCWPLWMPFQKPNTEIRQYCNHWRNFADIDDSWKSIIK 240
 DB 181 ADGYKMSLALNRTGRSIVVSCWPLWMPFQKPNTEIRQYCNHWRNFADIDDSWKSIIK 240
 QY 241 SILDWTFSNQERIVDVAGPGGNDPDMLVIGNFGLSNQOVTQMALWAINAAPLFMSNDL 300
 DB 241 SILDWTFSNQERIVDVAGPGGNDPDMLVIGNFGLSNQOVTQMALWAINAAPLFMSNDL 300
 QY 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRGDNFEVWERPLSLGLAWAVAMINRQEI 360
 DB 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRGDNFEVWERPLSLGLAWAVAMINRQEI 360
 QY 361 GPRSYTIAVASLGKGVACNACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
 DB 361 GPRSYTIAVASLGKGVACNACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 3
 ADJ88281
 ID ADJ88281 standard; protein; 417 AA.
 XX
 AC ADJ88281;
 XX
 DT 06-MAY-2004 (first entry)
 XX
 XX Human WT rGAL-12 (galactosidase).
 XX
 KW Galactosidase; GAL; gene therapy; lysosomal storage disease;
 KW Fabry's disease; Gaucher's disease; human; enzyme.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 400..417
 FT /note= "Encoded by ACT"
 XX
 XX US2004016021-A1.
 XX
 XX 22-JAN-2004.
 PD
 XX


```

PF 23-JUN-2003; 2003US-00602219.
XX
XX 26-FEB-1988; 88US-00160766.
XX 26-FEB-1988; 88US-00160771.
XX 15-JUL-1988; 88US-00219279.
XX 17-FEB-1989; 89US-00310881.
XX 05-MAY-1989; 89US-00347637.
XX 08-JUN-1989; 89US-00363138.
XX 22-OCT-1990; 90US-00600244.
XX 16-JAN-1991; 91US-00641617.
XX 26-JUL-1991; 91US-00737899.
XX 01-AUG-1991; 91US-00739143.
XX 31-JUL-1992; 92US-00923692.
XX 30-DEC-1992; 92US-00997733.
XX 29-DEC-1993; 93US-00176414.
XX 19-JAN-1994; 94US-00184237.
XX 14-OCT-1994; 94US-00324003.
XX 21-MAY-1999; 99US-00316572.
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.
XX
XX (TURP/) TURPEN T H.
XX (POGU/) FOGUE G P.
XX (ERWI/) ERWIN R L.
XX (GRIL/) GRILL L K.
XX
XX Turpen TH, Pogue GP, Erwin RL, Grill LK;
XX
XX WPI; 2004-108227/11.
XX N-PSDB; ADJ88280.
XX
XX New lysosomal enzymes, useful in treating human and animal lysosomal
XX storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
XX Claim 7; SEQ ID NO 16; 71pp; English.
XX
XX The invention relates to nucleotide encoding galactosidase (GAL). The
XX invention is useful in gene therapy. The polynucleotides and polypeptides
XX are useful in treating human and animal lysosomal storage diseases, e.g.
XX Fabry's disease and Gaucher's diseases. The present sequence is human GAL
XX protein.
XX
XX Sequence 417 AA;
XX
XX Query Match 100.0%; Score 2293; DB 8; Length 417;
XX Best Local Similarity 100.0%; Pred. No. 5.3e-220;
XX Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Qy 1 MOLRNPELHLCALALRFLALVSDIPGARALDNLARTPTMGWLHWRFCNLDQDEP 60
Db 1 MOLRNPELHLCALALRFLALVSDIPGARALDNLARTPTMGWLHWRFCNLDQDEP 60
Qy 61 DSCISEKLFMEAEMLVSEGHKADGYEYLICDDCWAPQDSEGRLOADPQRFPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGHKADGYEYLICDDCWAPQDSEGRLOADPQRFPHGIRQL 120
Qy 121 ANYVHSGKLGIVADVGNKTCAGPFGSGFYDDAOTFADWGDVLLKFDGCYCDLSLENL 180
Db 121 ANYVHSGKLGKIGIIVADVGNKTCAGPFGSGFYDDAOTFADWGDVLLKFDGCYCDLSLENL 180
Qy 181 ADGKHSLSALNRTGRSIVYSCWPLVYMPFKPNTEIRQYCNHWRNFADIDDSWKSIX 240
Db 181 ADGKHSLSALNRTGRSIVYSCWPLVYMPFKPNTEIRQYCNHWRNFADIDDSWKSIX 240
Qy 241 SILDWTSTNQRIVDVAGPGWNPDDMLVIGNFGLSNQVQVQVQVQVQVQVQVQVQVQV 300
Db 241 SILDWTSTNQRIVDVAGPGWNPDDMLVIGNFGLSNQVQVQVQVQVQVQVQVQVQVQV 300
Qy 301 RHISFOAKALLQDKDVIAINODPLGKQGYQLRQGDNPFWERPLSGLAWAVAMINRQBIG 360
Db 301 RHISFOAKALLQDKDVIAINODPLGKQGYQLRQGDNPFWERPLSGLAWAVAMINRQBIG 360
Qy 361 GPRSYTTIIVASLGKGVACNPACFITQLLPVKRKLGYEWTSLRSHINPTGTVILLQL 417

```

```

Db 361 GPRSYTTIIVASLGKGVACNPACFITQLLPVKRKLGYEWTSLRSHINPTGTVILLQL 417
XX
XX RESULT 4
XX ADM48685
XX ID ADM48685 standard; protein; 417 AA.
XX AC ADM48685;
XX XX
XX 03-JUN-2004 (first entry)
XX
XX Human wild type rGAL-12 protein.
XX
XX Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
XX lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
XX Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
XX enzyme.
XX
XX Homo sapiens.
XX
XX US2004023281-A1.
XX
XX 05-FEB-2004.
XX
XX 23-JUN-2003; 2003US-00602220.
XX
XX 26-FEB-1988; 88US-00160766.
XX 26-FEB-1988; 88US-00160771.
XX 17-FEB-1989; 89US-00310881.
XX 22-OCT-1990; 90US-00600244.
XX 31-JUL-1992; 92US-00923692.
XX 30-DEC-1992; 92US-00997733.
XX 29-DEC-1993; 93US-00176414.
XX 19-JAN-1994; 94US-00184237.
XX 14-OCT-1994; 94US-00324003.
XX 21-MAY-1999; 99US-00316572.
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.
XX
XX (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX (POGU/) FOGUE G P.
XX (ERWI/) ERWIN R L.
XX (GRIL/) GRILL L K.
XX
XX Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
XX
XX WPI; 2004-142650/14.
XX N-PSDB; ADM48684.
XX
XX New alpha-galactosidase polypeptides, useful in producing recombinant
XX lysosomal enzymes for the treatment of lysosomal storage diseases, such
XX as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
XX disease.
XX
XX Disclosure; SEQ ID NO 16; 72pp; English.
XX
XX The present invention relates to novel galactosidase (Gal) proteins such
XX as rGAL-12, rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions
XX of the present invention are useful for producing recombinant lysosomal
XX enzymes for enzyme replacement therapy for treating human and animal
XX lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
XX disease, Fabry disease and Tay-Sachs disease. The present sequence is
XX human wild type rGAL-12 protein used in the exemplification of the
XX invention.
XX
XX Sequence 417 AA;
XX
XX Query Match 100.0%; Score 2293; DB 8; Length 417;
XX Best Local Similarity 100.0%; Pred. No. 5.3e-220;
XX Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MOLNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMLDCEEP 60
 |||||
 Db 1 MOLNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMLDCEEP 60
 |||||
 QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQRDSEGRLOADPQRPFGIRQL 120
 |||||
 Db 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQRDSEGRLOADPQRPFGIRQL 120
 |||||
 QY 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
 |||||
 Db 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
 |||||
 QY 181 ADGKMSLALNRTGRSIVYSCWPLYWPFQKNTYIROYCNHWRNFADIDDSWKSIIK 240
 |||||
 Db 181 ADGKMSLALNRTGRSIVYSCWPLYWPFQKNTYIROYCNHWRNFADIDDSWKSIIK 240
 |||||
 QY 241 SILDWTSEFNOERIVDVAGPGGWNDDPMLVIGNFGLSWNQVQTOMALWAINAAPLFMSNDL 300
 |||||
 Db 241 SILDWTSEFNOERIVDVAGPGGWNDDPMLVIGNFGLSWNQVQTOMALWAINAAPLFMSNDL 300
 |||||
 QY 301 RHISPOKALLQDKVDVIAINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQIEG 360
 |||||
 Db 301 RHISPOKALLQDKVDVIAINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQIEG 360
 |||||
 QY 361 GPRSYTTIYAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLSRSHINPTGTLLQL 417
 |||||
 Db 361 GPRSYTTIYAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLSRSHINPTGTLLQL 417
 |||||

RESULT 5

ADU66920
 ID ADU66920 standard; protein; 417 AA.

XX AC ADU66920;

XX DT 10-FEB-2005 (first entry)

XX DE Human alpha-galactosidase protein #7.

XX KW Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase;
 enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;
 Niemann-Pick disease; Fabry's disease; Tay-Sachs disease;
 Hurler's syndrome; Hurler-Scheie syndrome; nephrotropic; human.

XX OS Homo sapiens.

XX PN US2004234516-A1.

XX PD 25-NOV-2004.

XX PF 21-MAY-2004; 2004US-00851388.

XX PR 26-JUL-2000; 2000US-00626127.

XX PR 13-NOV-2001; 2001US-00993059.

XX PR 20-MAR-2002; 2002US-00103327.

XX PA (LARG-) LARGE SCALE BIOLOGY CORP.

XX PI Garger SJ, Turpen TH, Kumagai MH;

XX DR WPI; 2004-821274/81.

XX DR N-PSDB; ADU66919.

XX PT A pharmaceutical composition comprising a lysosomal enzyme, useful for
 enzyme replacement therapy for the treatment of lysosomal storage
 diseases, such as Fabry's disease.

XX PS Disclosure; SEQ ID NO 16; 88pp; English.

XX CC The present invention relates to the production of human and animal
 lysosomal enzymes in plants by a transient plant expression system. The
 invention relates to glucocerebrosidase (GCB, GCR) and alpha-
 galactosidase enzymes having a post-translational modification provided

CC by the plant expression system. The invention is useful in enzyme
 CC replacement therapy for treating lysosomal storage diseases such as
 CC Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs
 CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
 CC also useful in researches for developing new approaches to medical
 CC treatment of lysosomal storage diseases and in industrial processes
 CC involving enzymatic substrate hydrolysis. The present sequence is the
 CC human alpha-galactosidase protein.

XX SQ Sequence 417 AA;

Query Match 100.0%; Score 2293; DB 8; Length 417;

Best Local Similarity 100.0%; Pred. No. 5.3e-220;

Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMLDCEEP 60

Db 1 MOLNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMLDCEEP 60

QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQRDSEGRLOADPQRPFGIRQL 120

Db 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAQRDSEGRLOADPQRPFGIRQL 120

QY 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180

Db 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180

QY 181 ADGKMSLALNRTGRSIVYSCWPLYWPFQKNTYIROYCNHWRNFADIDDSWKSIIK 240

Db 181 ADGKMSLALNRTGRSIVYSCWPLYWPFQKNTYIROYCNHWRNFADIDDSWKSIIK 240

QY 241 SILDWTSEFNOERIVDVAGPGGWNDDPMLVIGNFGLSWNQVQTOMALWAINAAPLFMSNDL 300

Db 241 SILDWTSEFNOERIVDVAGPGGWNDDPMLVIGNFGLSWNQVQTOMALWAINAAPLFMSNDL 300

QY 301 RHISPOKALLQDKVDVIAINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQIEG 360

Db 301 RHISPOKALLQDKVDVIAINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQIEG 360

QY 361 GPRSYTTIYAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLSRSHINPTGTLLQL 417

Db 361 GPRSYTTIYAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLSRSHINPTGTLLQL 417

RESULT 6

AEA27449

ID AEA27449 standard; protein; 417 AA.

XX AC AEA27449;

XX DT 11-AUG-2005 (first entry)

XX DE Human alpha-galactosidase protein, rGAL-12, SEQ ID NO: 16.

XX DE Gauchers disease; metabolic; neurologic disease; niemann pick disease;
 KW genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
 KW antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase;
 KW enzyme.

XX OS Homo sapiens.

XX PN US2005125859-A1.

XX PD 09-JUN-2005.

XX PF 08-NOV-2004; 2004US-00984389.

XX PR 26-JUL-2000; 2000US-00626127.

XX PR 13-NOV-2001; 2001US-00993059.

XX PR 20-MAR-2002; 2002US-00103327.

XX PA (LARG-) LARGE SCALE BIOLOGY CORP.

```

PI Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2005-404004/41.
DR N-PSDB; AEA2744B.
XX
PT New isolated polypeptides useful for producing lysosomal enzymes in
PT plants to be utilized in enzyme replacement therapy or for the
PT therapeutic treatment of human or animal lysosomal storage diseases, e.g.
PT Gaucher's disease.
XX
XX Disclosure; SEQ ID NO 16; 88pp; English.
XX
XX The present invention relates to the production of human and animal
XX lysosomal enzymes in plants by a transient plant expression system. The
XX invention relates to glucocerebrosidase (GCB, GCR) and alpha-
XX galactosidase (Gal) enzymes having a post-translational modification
XX provided by the plant expression system. The invention is useful in
XX enzyme replacement therapy for treating lysosomal storage diseases such
XX as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
XX disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
XX also useful in researches for developing new approaches to medical
XX treatment of lysosomal storage diseases and in industrial processes
XX involving enzymatic substrate hydrolysis. The present sequence is the
XX human alpha-galactosidase protein.
XX
SQ Sequence 417 AA;

Query Match 100.0%; Score 2293; DB 9; Length 417;
Best Local Similarity 100.0%; Pred. No. 5.3e-220; Indels 0; Gaps 0;
Matches 417; Conservative 0; Mismatches 0;

QY 1 MQLRNPELHGCALALRFLALVSDIPGARALDNGLARTPTMGLHWRFCNLDCCOEP 60
DB 1 MQLRNPELHGCALALRFLALVSDIPGARALDNGLARTPTMGLHWRFCNLDCCOEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKDAGYEYLCIDDCWAPQDSEGRLOADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKDAGYEYLCIDDCWAPQDSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSGKLGLGIYADVGNKTCAGFGSGFYDIDAQTFADMGVDLLKFDGCGYCDLSLENL 180
DB 121 ANYVHSGKLGLGIYADVGNKTCAGFGSGFYDIDAQTFADMGVDLLKFDGCGYCDLSLENL 180
QY 181 ADGKYHMSLALNRTGRSIVVSCWPLVWPFQKPNYTEIRQYCNHWNFAIDDDSWKSIK 240
DB 181 ADGKYHMSLALNRTGRSIVVSCWPLVWPFQKPNYTEIRQYCNHWNFAIDDDSWKSIK 240
QY 241 SILDWTSFNOERIVDVAGPGGWNDDPMLVGNFGLSNQVQTQWALWAIWAAPLFMSNDL 300
DB 241 SILDWTSFNOERIVDVAGPGGWNDDPMLVGNFGLSNQVQTQWALWAIWAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRGDNEFEVWERPLSLGLAWAVAMINROEIG 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRGDNEFEVWERPLSLGLAWAVAMINROEIG 360
QY 361 GPRSYTIAVASLKGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTLLQL 417
DB 361 GPRSYTIAVASLKGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTLLQL 417

RESULT 7
AAE28210
ID AAE28210 standard; protein; 421 AA.
XX
XX AAE28210;
AC
XX
XX 27-DEC-2002 (first entry)
DT
XX
XX Human rGAL-8 protein.
DE
XX
XX Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
KW therapeutic; rGAL-8.
XX

```

```

OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Cleavage-site 401..402
FT /note="CTPP cleavage site"
XX
XX US2002088024-A1.
PN
XX
XX 04-JUL-2002.
PD
XX
XX 13-NOV-2001; 2001US-00993059.
PF
XX
XX 26-JUL-2000; 2000US-00626127.
PR
XX
XX (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
PI
XX WPI; 2002-681656/73.
DR N-PSDB; AAD45223.
XX
XX Novel human alpha-galactosidase polypeptide useful for treating lysosomal
XX storage diseases.
PT
XX
XX Claim 7; Page 44; 88pp; English.
XX
XX The invention relates to human alpha-galactosidase truncated at the
XX carboxy terminus and the production of enzymatically active recombinant
XX human and animal lysosomal enzymes. The invention is useful for producing
XX lysosomal enzymes for treating lysosomal storage diseases, producing
XX altered or mutated proteins, enzymatically active or otherwise, to serve
XX as precursors or substrates for further in vivo or in vitro processing to
XX a specialised industrial form for research or therapeutic uses, to
XX produce more effective therapeutic enzyme, for producing antibodies
XX against lysosomal enzymes for medical diagnostic use, and in any
XX commercial process that involves substrate hydrolysis. The present
XX sequence is human rGAL-8 protein
XX
XX Sequence 421 AA;

Query Match 100.0%; Score 2293; DB 5; Length 421;
Best Local Similarity 100.0%; Pred. No. 5.4e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPELHGCALALRFLALVSDIPGARALDNGLARTPTMGLHWRFCNLDCCOEP 60
DB 1 MQLRNPELHGCALALRFLALVSDIPGARALDNGLARTPTMGLHWRFCNLDCCOEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKDAGYEYLCIDDCWAPQDSEGRLOADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKDAGYEYLCIDDCWAPQDSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSGKLGLGIYADVGNKTCAGFGSGFYDIDAQTFADMGVDLLKFDGCGYCDLSLENL 180
DB 121 ANYVHSGKLGLGIYADVGNKTCAGFGSGFYDIDAQTFADMGVDLLKFDGCGYCDLSLENL 180
QY 181 ADGKYHMSLALNRTGRSIVVSCWPLVWPFQKPNYTEIRQYCNHWNFAIDDDSWKSIK 240
DB 181 ADGKYHMSLALNRTGRSIVVSCWPLVWPFQKPNYTEIRQYCNHWNFAIDDDSWKSIK 240
QY 241 SILDWTSFNOERIVDVAGPGGWNDDPMLVGNFGLSNQVQTQWALWAIWAAPLFMSNDL 300
DB 241 SILDWTSFNOERIVDVAGPGGWNDDPMLVGNFGLSNQVQTQWALWAIWAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRGDNEFEVWERPLSLGLAWAVAMINROEIG 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRGDNEFEVWERPLSLGLAWAVAMINROEIG 360
QY 361 GPRSYTIAVASLKGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTLLQL 417
DB 361 GPRSYTIAVASLKGKGVACNCPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTLLQL 417

```

```
RESULT 8
ADD84751
ID   ADD84751 standard; protein; 421 AA.
XX
XX   ADD84751;
XX
XX   29-JAN-2004 (first entry)
XX
XX   Human alpha-galactosidase rGAL-8 polypeptide.
DE
XX
XX   Human; alpha-galactosidase; rGAL-4; lysosomal enzyme;
KW   enzyme replacement therapy; lysosomal disease; enzyme.
XX
XX   Homo sapiens.
XX
XX   US2003106095-A1.
XX
XX   05-JUN-2003.
XX
XX   20-MAR-2002; 2002US-00103327.
XX
XX   26-JUL-2000; 2000US-00626127.
XX
XX   13-NOV-2001; 2001US-00993059.
XX
XX   (GARG/) GARGER S J.
PA   (TURP/) TURPEN T H.
PA   (KUMA/) KUMAGAI M H.
XX
XX   Garger SJ, Turpen TH, Kumagai MH;
PI
XX
XX   WPI; 2003-801257/75.
DR
XX
XX   N-PSDB; ADD84750.
DR
XX
XX   New polynucleotide for producing active recombinant human and animal
PT   lysosomal enzymes in a plant expression system that can be used in enzyme
PT   replacement therapy.
XX
XX   Claim 7; SEQ ID NO 12; 77pp; English.
XX
XX   The invention relates to human alpha-galactosidase derivatives and the
CC   nucleic acids encoding them. The polypeptides are used in a method for
CC   producing active recombinant human and animal lysosomal enzymes in a
CC   plant expression system. The enzymes can be used in enzyme replacement
CC   therapy for the therapeutic treatment of human and animal lysosomal
CC   diseases. This sequence represents a human alpha-galactosidase derivative
CC   polypeptide of the invention.
XX
XX   Sequence 421 AA;
XX
XX   Query Match 100.0%; Score 2293; DB 7; Length 421;
XX   Best Local Similarity 100.0%; Pred. No. 5.4e-220;
XX   Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX   1 MOLNPELHGCALALRFLALVSDIFGARALDNGLARTPTMGWLHWFMCNLDCEEP 60
XX   1 MOLNPELHGCALALRFLALVSDIFGARALDNGLARTPTMGWLHWFMCNLDCEEP 60
XX
XX   61 DSCISEKLFMEAEMLVSEGWKQAGYEYLCITDCCWMAPOQDSEGRQLQADQRPFHGIRQL 120
XX   61 DSCISEKLFMEAEMLVSEGWKQAGYEYLCITDCCWMAPOQDSEGRQLQADQRPFHGIRQL 120
XX
XX   121 ANYVHSGKLGIYADVGNKTCAGFPGSFGYDIDAQTFADMGVDLLKFGDCYCDLSLENL 180
XX   121 ANYVHSGKLGIYADVGNKTCAGFPGSFGYDIDAQTFADMGVDLLKFGDCYCDLSLENL 180
XX
XX   181 ADGKYHMSLALNRGRISVYSCWPLVWBPQKENYTEIRQYCNHWRNFADIDDSWKSIIK 240
XX   181 ADGKYHMSLALNRGRISVYSCWPLVWBPQKENYTEIRQYCNHWRNFADIDDSWKSIIK 240
XX
XX   241 SILDWTSFNOERIVDVAGPGGWNDDPMLVIGNFGLSNWQQVTOQWALWAIMAAPLFMSNDL 300
XX   SILDWTSFNOERIVDVAGPGGWNDDPMLVIGNFGLSNWQQVTOQWALWAIMAAPLFMSNDL 300

Db 241 SILDWTSFNOERIVDVAGPGGWNDDPMLVIGNFGLSNWQQVTOQWALWAIMAAPLFMSNDL 300
Qy 301 RHISPOAKALLQDKDVIAINQDPLGKQGYOLROGDNFEVWERPLSGLAWAVAMINROEIG 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYOLROGDNFEVWERPLSGLAWAVAMINROEIG 360
Qy 361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
Db 361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 9
ADJ88277
ID   ADJ88277 standard; protein; 421 AA.
XX
XX   ADJ88277;
XX
XX   06-MAY-2004 (first entry)
XX
XX   Human WT rGAL-8 (galactosidase).
DE
XX
XX   Galactosidase; GAL; gene therapy; lysosomal storage disease;
KW   Fabry's disease; Gaucher's disease; human; enzyme.
XX
XX   Homo sapiens.
XX
XX   US2004016021-A1.
XX
XX   22-JAN-2004.
XX
XX   23-JUN-2003; 2003US-00602219.
XX
XX   26-FEB-1988; 88US-00160766.
XX
XX   26-FEB-1988; 88US-00160771.
XX
XX   15-JUL-1988; 88US-00219279.
XX
XX   17-FEB-1989; 89US-00310881.
XX
XX   05-MAY-1989; 89US-00347637.
XX
XX   08-JUN-1989; 89US-00363138.
XX
XX   22-OCT-1990; 90US-00600244.
XX
XX   16-JAN-1991; 91US-00641617.
XX
XX   26-JUL-1991; 91US-00737899.
XX
XX   01-AUG-1991; 91US-00739143.
XX
XX   31-JUL-1992; 92US-00923692.
XX
XX   30-DEC-1992; 92US-00997733.
XX
XX   29-DEC-1993; 93US-00176414.
XX
XX   19-JAN-1994; 94US-00184237.
XX
XX   14-OCT-1994; 94US-00324003.
XX
XX   21-MAY-1999; 99US-00316572.
XX
XX   26-JUL-2000; 2000US-00626127.
XX
XX   13-NOV-2001; 2001US-00993059.
XX
XX   (TURP/) TURPEN T H.
PA   (FOGU/) FOGUE G P.
PA   (ERWI/) ERWIN R L.
PA   (GRIL/) GRILL L K.
XX
XX   Turpen TH, Pogue GP, Erwin RL, Grill LX;
XX
XX   WPI; 2004-108227/11.
XX
XX   N-PSDB; ADJ88276.
XX
XX   New lysosomal enzymes, useful in treating human and animal lysosomal
PT   storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
XX   Claim 7; SEQ ID NO 12; 71pp; English.
XX
XX   The invention relates to nucleotide encoding galactosidase (GAL). The
CC   invention is useful in gene therapy. The polynucleotides and polypeptides
CC   are useful in treating human and animal lysosomal storage diseases, e.g.
CC   Fabry's disease and Gaucher's diseases. The present sequence is human GAL
CC   protein.
XX
XX   Sequence 421 AA;
XX
```

Query Match 100.0%; Score 2293; DB 8; Length 421;
 Best Local Similarity 100.0%; Pred. No. 5.4e-220;
 Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPELHGGCALALRFLALVSDIFGARALDNGLARTPTMGWLHWRFCNLDCCOEP 60
 DB 1 MQLRNPELHGGCALALRFLALVSDIFGARALDNGLARTPTMGWLHWRFCNLDCCOEP 60

QY 61 DSCISEKLFMEALMVSEGWKAGYEYLCIDDCWMAPOQDSGRQLQADQRPFGHGIQL 120
 DB 61 DSCISEKLFMEALMVSEGWKAGYEYLCIDDCWMAPOQDSGRQLQADQRPFGHGIQL 120

QY 121 ANYVHSGKLGKIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
 DB 121 ANYVHSGKLGKIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180

QY 181 ADGKMSLALNRTGRSIVYSCWPLYMWPFPKPNTEIRQYCNHWRNFADIDDSWKSIIK 240
 DB 181 ADGKMSLALNRTGRSIVYSCWPLYMWPFPKPNTEIRQYCNHWRNFADIDDSWKSIIK 240

QY 241 SILDWTFSNQRERIVDVAGPGGNDPDMVLVGNFGLSNQOVTQWALWAIAPLFMSNDL 300
 DB 241 SILDWTFSNQRERIVDVAGPGGNDPDMVLVGNFGLSNQOVTQWALWAIAPLFMSNDL 300

QY 301 RHISPOKALLQDKVDIAINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQEI 360
 DB 301 RHISPOKALLQDKVDIAINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQEI 360

QY 361 GPRSYYTIAVASLGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTLLQL 417
 DB 361 GPRSYYTIAVASLGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTLLQL 417

RESULT 10
 ADW48681
 ID ADW48681 standard; protein; 421 AA.
 XX AC ADW48681;
 DT 03-JUN-2004 (first entry)
 XX Human wild type rGAL-8 protein.
 DE Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
 XX lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
 KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
 KW enzyme.
 XX Homo sapiens.
 OS US2004023281-A1.
 XX 05-FEB-2004.
 XX 23-JUN-2003; 2003US-00602220.
 XX 26-FEB-1988; 88US-00160766.
 XX 26-FEB-1988; 88US-00160771.
 XX 17-FEB-1989; 89US-00310881.
 XX 22-OCT-1990; 90US-00600244.
 XX 31-JUL-1992; 92US-00923692.
 XX 30-DEC-1992; 92US-00997733.
 XX 29-DEC-1993; 93US-00176414.
 XX 19-JAN-1994; 94US-00184237.
 XX 14-OCT-1994; 94US-00324003.
 XX 21-MAY-1999; 99US-00316572.
 XX 26-JUL-2000; 2000US-00626127.
 XX 13-NOV-2001; 2001US-00993059.
 XX (TURP/) TURPEN T H.
 PA (KUMA/) KUMAGAI M H.
 PA (POGU/) POGUE G P.

PA (ERWI/) ERWIN R L.
 PA (GRIL/) GRILL L K.
 XX Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
 XX WPI: 2004-142650/14.
 DR N-PSDB; ADM48680.
 XX New alpha-galactosidase polypeptides, useful in producing recombinant
 PT lysosomal enzymes for the treatment of lysosomal storage diseases, such
 PT as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
 PT disease.
 XX Disclosure; SEQ ID NO 12; 72pp; English.
 XX The present invention relates to novel galactosidase (Gal) proteins such
 CC as rGAL-12, rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions
 CC of the present invention are useful for producing recombinant lysosomal
 CC enzymes for enzyme replacement therapy for treating human and animal
 CC lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
 CC disease, Fabry disease and Tay-Sachs disease. The present sequence is
 CC human wild type rGAL-8 protein used in the exemplification of the
 CC invention.
 XX Sequence 421 AA;
 SQ

Query Match 100.0%; Score 2293; DB 8; Length 421;
 Best Local Similarity 100.0%; Pred. No. 5.4e-220;
 Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPELHGGCALALRFLALVSDIFGARALDNGLARTPTMGWLHWRFCNLDCCOEP 60
 DB 1 MQLRNPELHGGCALALRFLALVSDIFGARALDNGLARTPTMGWLHWRFCNLDCCOEP 60

QY 61 DSCISEKLFMEALMVSEGWKAGYEYLCIDDCWMAPOQDSGRQLQADQRPFGHGIQL 120
 DB 61 DSCISEKLFMEALMVSEGWKAGYEYLCIDDCWMAPOQDSGRQLQADQRPFGHGIQL 120

QY 121 ANYVHSGKLGKIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
 DB 121 ANYVHSGKLGKIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180

QY 181 ADGKMSLALNRTGRSIVYSCWPLYMWPFPKPNTEIRQYCNHWRNFADIDDSWKSIIK 240
 DB 181 ADGKMSLALNRTGRSIVYSCWPLYMWPFPKPNTEIRQYCNHWRNFADIDDSWKSIIK 240

QY 241 SILDWTFSNQRERIVDVAGPGGNDPDMVLVGNFGLSNQOVTQWALWAIAPLFMSNDL 300
 DB 241 SILDWTFSNQRERIVDVAGPGGNDPDMVLVGNFGLSNQOVTQWALWAIAPLFMSNDL 300

QY 301 RHISPOKALLQDKVDIAINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQEI 360
 DB 301 RHISPOKALLQDKVDIAINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQEI 360

QY 361 GPRSYYTIAVASLGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTLLQL 417
 DB 361 GPRSYYTIAVASLGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTLLQL 417

RESULT 11
 ADU66916
 ID ADU66916 standard; protein; 421 AA.
 XX AC ADU66916;
 XX 10-FEB-2005 (first entry)
 DE Human alpha-galactosidase protein #5.
 XX Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase;
 KW enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;
 KW Niemann-Pick disease; Fabry's disease; Tay-Sachs disease; Hurler's
 KW Hurler's syndrome; Hurler-Scheele syndrome; nephrotrophic; human.

```
XX OS Homo sapiens.
XX PN US2004234516-A1.
XX PD 25-NOV-2004.
XX PF 21-MAY-2004; 2004US-00851388.
XX PR 26-JUL-2000; 2000US-00626127.
XX PR 13-NOV-2001; 2001US-00993059.
XX PR 20-MAR-2002; 2002US-00103327.
XX PA (LARG-) LARGE SCALE BIOLOGY CORP.
XX PI Garger SJ, Turpen TH, Kumagai MH;
XX DR WPI; 2004-821274/81.
XX DR N-PSDB; ADU66915.
XX PT A pharmaceutical composition comprising a lysosomal enzyme, useful for
XX PT enzyme replacement therapy for the treatment of lysosomal storage
XX PT diseases, such as Fabry's disease.
XX PS Disclosure; SEQ ID NO 12; 88pp; English.
XX CC The present invention relates to the production of human and animal
XX CC lysosomal enzymes in plants by a transient plant expression system. The
XX CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
XX CC galactosidase enzymes having a post-translational modification provided
XX CC by the plant expression system. The invention is useful in enzyme
XX CC replacement therapy for treating lysosomal storage diseases such as
XX CC Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs
XX CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
XX CC also useful in researches for developing new approaches to medical
XX CC treatment of lysosomal storage diseases and in industrial processes
XX CC involving enzymatic substrate hydrolysis. The present sequence is the
XX CC human alpha-galactosidase protein.
XX SQ Sequence 421 AA;
Query Match 100.0%; Score 2293; DB 8; Length 421;
Best Local Similarity 100.0%; Pred. No. 5.4e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MQLRNPDLHLCALALRFLALVSWDIPGARALDNLGLARTMTGWLHWRFCMLDCQEEP 60
DB 1 MQLRNPDLHLCALALRFLALVSWDIPGARALDNLGLARTMTGWLHWRFCMLDCQEEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKDAGYEYLCIDCWMAPOQDSGRLOADPQRPFGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKDAGYEYLCIDCWMAPOQDSGRLOADPQRPFGIRQL 120
QY 121 ANYVHSGKLGLGIYADVGNKTCAGFPGSFGYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
DB 121 ANYVHSGKLGLGIYADVGNKTCAGFPGSFGYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
QY 181 ADGKXMSLALNTRGRISVSCWPLVWMPFQPNYTEIRQYCNHWNFNADIDDSWKSIIK 240
DB 181 ADGKXMSLALNTRGRISVSCWPLVWMPFQPNYTEIRQYCNHWNFNADIDDSWKSIIK 240
QY 241 SILDWTSFNQRIVDVNAGPGGNDPDMLVINGFGLSNQVQTQWALWAINAAPIFWSNDL 300
DB 241 SILDWTSFNQRIVDVNAGPGGNDPDMLVINGFGLSNQVQTQWALWAINAAPIFWSNDL 300
QY 301 RHISPOKALLQDKVTAIINQDPLGKQGYQLRQGDNFVWVERPLISGLAWAVAMINRQEIG 360
DB 301 RHISPOKALLQDKVTAIINQDPLGKQGYQLRQGDNFVWVERPLISGLAWAVAMINRQEIG 360
QY 361 GPRSYTTIATVSLGKGVACNPACFTITQLLPVKRLGIFYEWTSLRASHINPTGTLLQL 417
DB 361 GPRSYTTIATVSLGKGVACNPACFTITQLLPVKRLGIFYEWTSLRASHINPTGTLLQL 417
```

```
RESULT 12
AEA27445
ID AEA27445 standard; protein; 421 AA.
XX AC AEA27445;
XX AS AEA27445;
XX DT 11-AUG-2005 (first entry)
XX DE Human alpha-galactosidase protein, rGAL-8, SEQ ID NO: 12.
XX KW Gauchers disease; metabolic; neurological disease; niemann pick disease;
XX KW genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
XX KW antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase;
XX KW enzyme.
XX OS Homo sapiens.
XX PN US2005125859-A1.
XX PD 09-JUN-2005.
XX PF 08-NOV-2004; 2004US-00984389.
XX PR 26-JUL-2000; 2000US-00626127.
XX PR 13-NOV-2001; 2001US-00993059.
XX PR 20-MAR-2002; 2002US-00103327.
XX PA (LARG-) LARGE SCALE BIOLOGY CORP.
XX PI Garger SJ, Turpen TH, Kumagai MH;
XX DR WPI; 2005-404004/41.
XX DR N-PSDB; AEA27444.
XX PT New isolated polypeptides useful for producing lysosomal enzymes in
XX PT plants to be utilized in enzyme replacement therapy or for the
XX PT therapeutic treatment of human or animal lysosomal storage diseases, e.g.
XX PT Gaucher's disease.
XX FS Claim 28; SEQ ID NO 12; 88pp; English.
XX CC The present invention relates to the production of human and animal
XX CC lysosomal enzymes in plants by a transient plant expression system. The
XX CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
XX CC galactosidase (Gal) enzymes having a post-translational modification
XX CC provided by the plant expression system. The invention is useful in
XX CC enzyme replacement therapy for treating lysosomal storage diseases such
XX CC as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
XX CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
XX CC also useful in researches for developing new approaches to medical
XX CC treatment of lysosomal storage diseases and in industrial processes
XX CC involving enzymatic substrate hydrolysis. The present sequence is the
XX CC human alpha-galactosidase protein.
XX SQ Sequence 421 AA;
Query Match 100.0%; Score 2293; DB 9; Length 421;
Best Local Similarity 100.0%; Pred. No. 5.4e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MQLRNPDLHLCALALRFLALVSWDIPGARALDNLGLARTMTGWLHWRFCMLDCQEEP 60
DB 1 MQLRNPDLHLCALALRFLALVSWDIPGARALDNLGLARTMTGWLHWRFCMLDCQEEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKDAGYEYLCIDCWMAPOQDSGRLOADPQRPFGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKDAGYEYLCIDCWMAPOQDSGRLOADPQRPFGIRQL 120
QY 121 ANYVHSGKLGLGIYADVGNKTCAGFPGSFGYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
DB 121 ANYVHSGKLGLGIYADVGNKTCAGFPGSFGYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
```

Qy 181 ADGKXMSLALNRTGRSIVSCWPLVMPFQKPNYTEIRQYCNHWNRFADIDDSWKSIIK 240
Db 181 ADGKXMSLALNRTGRSIVSCWPLVMPFQKPNYTEIRQYCNHWNRFADIDDSWKSIIK 240
Qy 241 SILDWTSPNOERIVDVAGPGGWNDDMLVIGNFGLSWNQVQTQALWAIWAAPLFMSNDL 300
Db 241 SILDWTSPNOERIVDVAGPGGWNDDMLVIGNFGLSWNQVQTQALWAIWAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKDVIATINQDPLGKQGYQLRQGDNEFVWERPLSLGLAWAVAMINRQEI 360
Db 301 RHISPOKALLQDKDVIATINQDPLGKQGYQLRQGDNEFVWERPLSLGLAWAVAMINRQEI 360
Qy 361 GPRSYTTIIVASLKGKGVACNPFACITQLLPVKRKLGFYEWTSRLSHINPTGTLLQL 417
Db 361 GPRSYTTIIVASLKGKGVACNPFACITQLLPVKRKLGFYEWTSRLSHINPTGTLLQL 417
RESULT 13
AAE28213
XX AAE28213 standard; protein; 423 AA.
AC AAE28213;
XX
DT 27-DEC-2002 (first entry)
XX Human rGAL-12R protein.
XX Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
KW therapeutic; rGAL-12R.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Cleavage-site 401..402
FT /note= "CTPP cleavage site"
FT
XX US2002088024-A1.
XX
XX 04-JUL-2002.
XX
XX 13-NOV-2001; 2001US-00993059.
XX
XX 26-JUL-2000; 2000US-00626127.
XX
XX (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
XX
XX WPI; 2002-681656/73.
DR N-PSDB; AAD45226.
XX
XX Novel human alpha-galactosidase polypeptide useful for treating lysosomal
PT storage diseases.
XX
XX Claim 7; Page 51-52; 88pp; English.
XX
XX The invention relates to human alpha-galactosidase truncated at the
CC carboxy terminus and the production of enzymatically active recombinant
CC human and animal lysosomal enzymes. The invention is useful for producing
CC lysosomal enzymes for treating lysosomal storage diseases, producing
CC altered or mutated proteins, enzymatically active or otherwise, to serve
CC as precursors or substrates for further in vivo or in vitro processing to
CC a specialised industrial form for research or therapeutic uses, to
CC produce more effective therapeutic enzyme, for producing antibodies
CC against lysosomal enzymes for medical diagnostic use, and in any
CC commercial process that involves substrate hydrolysis. The present
CC sequence is human rGAL-12R protein
XX
XX Sequence 423 AA;

Query Match

100.0%; Score 2293; DB 5; Length 423;

Best Local Similarity 100.0%; Pred. No. 5.4e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MQLRNPDLHGLCALALRFLALVSWDIPCARALDNLARTPTMGLHWRFCNLDCCQEEP 60
Db 1 MQLRNPDLHGLCALALRFLALVSWDIPCARALDNLARTPTMGLHWRFCNLDCCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGMKADAGYEYLCIDDCWMAPOQSDSEGRLOADPQPFPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGMKADAGYEYLCIDDCWMAPOQSDSEGRLOADPQPFPHGIRQL 120
Qy 121 ANYVHSGKLGLGIYADVGNKTCAGFCGSGYDIDAQTFADGWYDILLKFDGCYCDLSLNL 180
Db 121 ANYVHSGKLGLGIYADVGNKTCAGFCGSGYDIDAQTFADGWYDILLKFDGCYCDLSLNL 180
Qy 181 ADGKXMSLALNRTGRSIVSCWPLVMPFQKPNYTEIRQYCNHWNRFADIDDSWKSIIK 240
Db 181 ADGKXMSLALNRTGRSIVSCWPLVMPFQKPNYTEIRQYCNHWNRFADIDDSWKSIIK 240
Qy 241 SILDWTSPNOERIVDVAGPGGWNDDMLVIGNFGLSWNQVQTQALWAIWAAPLFMSNDL 300
Db 241 SILDWTSPNOERIVDVAGPGGWNDDMLVIGNFGLSWNQVQTQALWAIWAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKDVIATINQDPLGKQGYQLRQGDNEFVWERPLSLGLAWAVAMINRQEI 360
Db 301 RHISPOKALLQDKDVIATINQDPLGKQGYQLRQGDNEFVWERPLSLGLAWAVAMINRQEI 360
Qy 361 GPRSYTTIIVASLKGKGVACNPFACITQLLPVKRKLGFYEWTSRLSHINPTGTLLQL 417
Db 361 GPRSYTTIIVASLKGKGVACNPFACITQLLPVKRKLGFYEWTSRLSHINPTGTLLQL 417
RESULT 14
ADD84757
ID ADD84757 standard; protein; 423 AA.
XX
AC ADD84757;
XX
DT 29-JAN-2004 (first entry)
XX
XX Human alpha-galactosidase rGAL-12R polypeptide.
XX
XX Human; alpha-galactosidase; rGAL-4; lysosomal enzyme;
KW enzyme replacement therapy; lysosomal disease; enzyme.
XX
XX Homo sapiens.
XX
XX US2003106095-A1.
XX
XX 05-JUN-2003.
XX
XX 20-MAR-2002; 2002US-00103327.
XX
XX 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX
XX (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
XX
XX WPI; 2003-801257/75.
DR N-PSDB; ADD84756.
XX
XX New polynucleotide for producing active recombinant human and animal
PT lysosomal enzymes in a plant expression system that can be used in enzyme
PT replacement therapy.
XX
XX Claim 7; SEQ ID NO 18; 77pp; English.
XX
XX The invention relates to human alpha-galactosidase derivatives and the
CC nucleic acids encoding them. The polypeptides are used in a method for

CC producing active recombinant human and animal lysosomal enzymes in a
CC plant expression system. The enzymes can be used in enzyme replacement
CC therapy for the therapeutic treatment of human and animal lysosomal
CC diseases. This sequence represents a human alpha-galactosidase derivative
CC polypeptide of the invention.
XX
SQ Sequence 423 AA;

Query Match 100.0%; Score 2293; DB 7; Length 423;
Best Local Similarity 100.0%; Pred. No. 5.4e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGLHWRFCMLDCCQEEP 60
DB 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGLHWRFCMLDCCQEEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAPOQDSGRLOADPQRPFGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAPOQDSGRLOADPQRPFGIRQL 120
QY 121 ANYVHSGKLGKIYADVGNKTCAGFPFGSGFYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
DB 121 ANYVHSGKLGKIYADVGNKTCAGFPFGSGFYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
QY 181 ADGKMSLALNRTGRSIVVSCWPLMPPFQKPNYTEIRQYCNHWNFNADIDDSWKSIIK 240
DB 181 ADGKMSLALNRTGRSIVVSCWPLMPPFQKPNYTEIRQYCNHWNFNADIDDSWKSIIK 240
QY 241 SILDWTSTFNOERIVDVAGPGGNDPDMVLVGNFGLSNQVQTQWALWAIWAAPLFMSNDL 300
DB 241 SILDWTSTFNOERIVDVAGPGGNDPDMVLVGNFGLSNQVQTQWALWAIWAAPLFMSNDL 300
QY 301 RHISPOKALLQDKVDVAINQDPLGKGYQLROGDNFEVWERPLSLGLAWAVAMINRQEIG 360
DB 301 RHISPOKALLQDKVDVAINQDPLGKGYQLROGDNFEVWERPLSLGLAWAVAMINRQEIG 360
QY 361 GPRSYTIAVSLGKGVACNACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
DB 361 GPRSYTIAVSLGKGVACNACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 15
ADJ88283
ID ADJ88283 standard; protein; 423 AA.
XX
AC ADJ88283;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human WT rGAL-12R (galactosidase).
XX
KW Galactosidase; GAL; gene therapy; lysosomal storage disease;
KW Fabry's disease; Gaucher's disease; human; enzyme.
XX
OS Homo sapiens.
XX
PN US2004016021-A1.
XX
PD 22-JAN-2004.
XX
PF 23-JUN-2003; 2003US-00602219.
XX
PR 26-FEB-1988; 88US-00160766.
PR 26-FEB-1988; 88US-00160771.
PR 15-JUL-1988; 88US-00219279.
PR 17-FEB-1989; 89US-00310881.
PR 05-MAY-1989; 89US-00347637.
PR 08-JUN-1989; 89US-00363138.
PR 22-OCT-1990; 90US-00600244.
PR 16-JAN-1991; 91US-00641617.
PR 26-JUL-1991; 91US-00737899.
PR 01-AUG-1991; 91US-00739143.
PR 31-JUL-1992; 92US-00923692.

PR 30-DEC-1992; 92US-00997733.
PR 29-DEC-1993; 93US-00176414.
PR 19-JAN-1994; 94US-00184237.
PR 14-OCT-1994; 94US-00324003.
PR 21-MAY-1999; 99US-00316572.
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX
PA (TURP/) TURPEN T H.
PA (POGU/) POGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX
XX Turpen TH, Pogue GP, Erwin RL, Grill LK;
DR WPI; 2004-108227/11.
XX N-PSDB; ADJ88282.
XX
XX New lysosomal enzymes, useful in treating human and animal lysosomal
PT storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
XX Claim 7; SEQ ID NO 18; 71pp; English.
XX
XX The invention relates to nucleotide encoding galactosidase (GAL). The
CC invention is useful in gene therapy. The polynucleotides and polypeptides
CC are useful in treating human and animal lysosomal storage diseases, e.g.
CC Fabry's disease and Gaucher's diseases. The present sequence is human GAL
CC protein.
XX
SQ Sequence 423 AA;

Query Match 100.0%; Score 2293; DB 8; Length 423;
Best Local Similarity 100.0%; Pred. No. 5.4e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGLHWRFCMLDCCQEEP 60
DB 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGLHWRFCMLDCCQEEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAPOQDSGRLOADPQRPFGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAPOQDSGRLOADPQRPFGIRQL 120
QY 121 ANYVHSGKLGKIYADVGNKTCAGFPFGSGFYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
DB 121 ANYVHSGKLGKIYADVGNKTCAGFPFGSGFYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
QY 181 ADGKMSLALNRTGRSIVVSCWPLMPPFQKPNYTEIRQYCNHWNFNADIDDSWKSIIK 240
DB 181 ADGKMSLALNRTGRSIVVSCWPLMPPFQKPNYTEIRQYCNHWNFNADIDDSWKSIIK 240
QY 241 SILDWTSTFNOERIVDVAGPGGNDPDMVLVGNFGLSNQVQTQWALWAIWAAPLFMSNDL 300
DB 241 SILDWTSTFNOERIVDVAGPGGNDPDMVLVGNFGLSNQVQTQWALWAIWAAPLFMSNDL 300
QY 301 RHISPOKALLQDKVDVAINQDPLGKGYQLROGDNFEVWERPLSLGLAWAVAMINRQEIG 360
DB 301 RHISPOKALLQDKVDVAINQDPLGKGYQLROGDNFEVWERPLSLGLAWAVAMINRQEIG 360
QY 361 GPRSYTIAVSLGKGVACNACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
DB 361 GPRSYTIAVSLGKGVACNACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

Search completed: December 31, 2005, 23:43:14
Job time : 155.421 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 23:36:26 ; Search time 43.1379 Seconds
(without alignments)
799.198 Million cell updates/sec

Title: US-10-602-220-16

Perfect score: 2293
Sequence: 1 MQLRNPGLHGCALALRFLA.....EWTSLRSHINPTGTLLQL 417

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
- 2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
- 3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
- 4: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.pep:*
- 5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2293	100.0	417	2	US-09-993-059-16
2	2293	100.0	417	2	US-10-103-327-16
3	2293	100.0	421	2	US-09-993-059-12
4	2293	100.0	421	2	US-10-103-327-12
5	2293	100.0	423	2	US-09-993-059-18
6	2293	100.0	423	2	US-10-103-327-18
7	2293	100.0	427	2	US-09-993-059-14
8	2293	100.0	427	2	US-10-103-327-14
9	2293	100.0	429	1	US-07-602-824-2
10	2293	100.0	429	1	US-07-602-608-3
11	2293	100.0	429	1	US-07-983-451-2
12	2293	100.0	429	1	US-08-261-578-3
13	2293	100.0	429	1	US-08-261-577-7
14	2293	100.0	429	1	US-08-261-577-9
15	2293	100.0	431	2	US-09-993-059-10
16	2293	100.0	431	2	US-10-103-327-10
17	2293	100.0	435	2	US-09-993-059-6
18	2293	100.0	435	2	US-10-103-327-6
19	2289	99.8	429	2	US-09-070-356-4
20	2288	99.8	424	2	US-09-993-059-8
21	2288	99.8	424	2	US-10-103-327-8
22	2288	99.8	428	2	US-09-993-059-4
23	2288	99.8	428	2	US-10-103-327-4
24	2275	99.2	428	6	5179023-4
25	2226	97.1	409	2	US-09-993-059-22
26	2226	97.1	409	2	US-10-103-327-22
27	2213	96.5	401	2	US-09-993-059-20

28	2213	96.5	401	2	US-10-103-327-20	Sequence 20, Appl
29	2131	92.9	386	2	US-09-176-666-11	Sequence 11, Appl
30	2131	92.9	387	2	US-09-176-666-10	Sequence 10, Appl
31	2131	92.9	388	2	US-09-176-666-9	Sequence 9, Appl
32	2131	92.9	389	2	US-09-176-666-8	Sequence 8, Appl
33	2131	92.9	390	2	US-09-176-666-7	Sequence 7, Appl
34	2131	92.9	391	2	US-09-176-666-6	Sequence 6, Appl
35	2131	92.9	392	2	US-09-176-666-5	Sequence 5, Appl
36	2131	92.9	393	2	US-09-176-666-4	Sequence 4, Appl
37	2131	92.9	394	2	US-09-176-666-3	Sequence 3, Appl
38	2131	92.9	396	2	US-09-176-666-2	Sequence 2, Appl
39	2131	92.9	398	2	US-08-928-881-26	Sequence 26, Appl
40	2131	92.9	398	2	US-09-176-666-1	Sequence 1, Appl
41	2131	92.9	398	2	US-09-543-921-26	Sequence 26, Appl
42	2131	92.9	398	2	US-09-266-014-4	Sequence 26, Appl
43	2131	92.9	398	2	US-09-491-759-26	Sequence 26, Appl
44	2131	92.9	398	2	US-10-360-101-202	Sequence 202, App
45	2110	92.0	381	2	US-09-176-666-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1

US-09-993-059-16
; Sequence 16, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-16

Query Match	100.0%;	Score 2293;	DB 2;	Length 417;
Best Local Similarity	100.0%;	Pred. No. 1.5e-232;		
Matches	417;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Qy	1	MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCNLDQCEP	60	
Db	1	MQLRNPGLHGCALALRFLALVSWDIPGARALDNLGLARTPTMGWLHWRFCNLDQCEP	60	
Qy	61	DSCTSEKLFMEALMVSEGMKADGYELCIDDCWMAPOQDSEGRLOADPQRFPHGIROL	120	
Db	61	DSCTSEKLFMEALMVSEGMKADGYELCIDDCWMAPOQDSEGRLOADPQRFPHGIROL	120	
Qy	121	ANYVHSGKLGLIYADVGNKTCAGFGSGFYDIDAQTFADMGVDLLKFDGCDLSLENL	180	
Db	121	ANYVHSGKLGLIYADVGNKTCAGFGSGFYDIDAQTFADMGVDLLKFDGCDLSLENL	180	
Qy	181	ADGKMSLALNRGSRISIVSCWPLYMFWFKPNYTEIRQYCNHWNFAIDDSWSKIK	240	
Db	181	ADGKMSLALNRGSRISIVSCWPLYMFWFKPNYTEIRQYCNHWNFAIDDSWSKIK	240	
Qy	241	SILDWTSFNQERIVDVAGPGWNDPDMVLVGNFGLSNQVQVOTMALWMAAPLPMNDL	300	
Db	241	SILDWTSFNQERIVDVAGPGWNDPDMVLVGNFGLSNQVQVOTMALWMAAPLPMNDL	300	
Qy	301	RHISPOKALLQDKVIAINQDPLGKQYQLRQGDNFVWVERPLSGLAWAVAMINRQIG	360	
Db	301	RHISPOKALLQDKVIAINQDPLGKQYQLRQGDNFVWVERPLSGLAWAVAMINRQIG	360	
Qy	361	GPRSYTIAVSLGKGVACNACPFITQLLLPVKRLGIFYEWTSLRSHINPTGTLLQL	417	

```
|||||
Db 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRLGFEYWTSLRSLHNPITGTLLQL 417

RESULT 2
US-10-103-327-16
; Sequence 16, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; CURRENT APPLICATION NUMBER: US/10/103,327
; FILE REFERENCE: 008010087CPUS06
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 417
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-103-327-16

Query Match 100.0%; Score 2293; DB 2; Length 417;
Best Local Similarity 100.0%; Pred. No. 1.5e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGLHWRFCNLDCEEP 60
Db 1 MQLRNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGLHWRFCNLDCEEP 60
Qy 61 DSCISEKLFMEALMVSEGWKADAGEYLCIDCWMAQRDSEGRLOADPQRPFGIRQL 120
Db 61 DSCISEKLFMEALMVSEGWKADAGEYLCIDCWMAQRDSEGRLOADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDDLKFDGCYCDL 180
Db 121 ANYVHSGKLGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDDLKFDGCYCDL 180
Qy 181 ADGKMSLALNRTGRSIVSCWPLYMFPQKPNYTEIRQYCNHWRNFADIDDSWKS 240
Db 181 ADGKMSLALNRTGRSIVSCWPLYMFPQKPNYTEIRQYCNHWRNFADIDDSWKS 240
Qy 241 SILDWTSFNQRIVDVAGPGWNPDMVLVGNFGLSNQOVTOMALWAIMAALPFW 300
Db 241 SILDWTSFNQRIVDVAGPGWNPDMVLVGNFGLSNQOVTOMALWAIMAALPFW 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKGYQLRQGDNFVWVERPLSGLAWAVAMINR 360
Db 301 RHISPOKALLQDKVIAINQDPLGKGYQLRQGDNFVWVERPLSGLAWAVAMINR 360
Qy 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRLGFEYWTSLRSLHNPITGT 417
Db 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRLGFEYWTSLRSLHNPITGT 417

RESULT 3
US-09-993-059-12
; Sequence 12, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; CURRENT APPLICATION NUMBER: US/09/993,059
; FILE REFERENCE: 008010087CPUS06
; CURRENT FILING DATE: 2001-11-13
```

```
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PR1
; ORGANISM: Homo sapiens
US-09-993-059-12

Query Match 100.0%; Score 2293; DB 2; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGLHWRFCNLDCEEP 60
Db 1 MQLRNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGLHWRFCNLDCEEP 60
Qy 61 DSCISEKLFMEALMVSEGWKADAGEYLCIDCWMAQRDSEGRLOADPQRPFGIRQL 120
Db 61 DSCISEKLFMEALMVSEGWKADAGEYLCIDCWMAQRDSEGRLOADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDDLKFDGCYCDL 180
Db 121 ANYVHSGKLGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDDLKFDGCYCDL 180
Qy 181 ADGKMSLALNRTGRSIVSCWPLYMFPQKPNYTEIRQYCNHWRNFADIDDSWKS 240
Db 181 ADGKMSLALNRTGRSIVSCWPLYMFPQKPNYTEIRQYCNHWRNFADIDDSWKS 240
Qy 241 SILDWTSFNQRIVDVAGPGWNPDMVLVGNFGLSNQOVTOMALWAIMAALPFW 300
Db 241 SILDWTSFNQRIVDVAGPGWNPDMVLVGNFGLSNQOVTOMALWAIMAALPFW 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKGYQLRQGDNFVWVERPLSGLAWAVAMINR 360
Db 301 RHISPOKALLQDKVIAINQDPLGKGYQLRQGDNFVWVERPLSGLAWAVAMINR 360
Qy 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRLGFEYWTSLRSLHNPITGT 417
Db 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRLGFEYWTSLRSLHNPITGT 417

RESULT 4
US-10-103-327-12
; Sequence 12, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; CURRENT APPLICATION NUMBER: US/10/103,327
; FILE REFERENCE: 008010087CPUS06
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-103-327-12

Query Match 100.0%; Score 2293; DB 2; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGLHWRFCNLDCEEP 60
Db 1 MQLRNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGLHWRFCNLDCEEP 60
Qy 61 DSCISEKLFMEALMVSEGWKADAGEYLCIDCWMAQRDSEGRLOADPQRPFGIRQL 120
Db 61 DSCISEKLFMEALMVSEGWKADAGEYLCIDCWMAQRDSEGRLOADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDDLKFDGCYCDL 180
Db 121 ANYVHSGKLGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDDLKFDGCYCDL 180
Qy 181 ADGKMSLALNRTGRSIVSCWPLYMFPQKPNYTEIRQYCNHWRNFADIDDSWKS 240
Db 181 ADGKMSLALNRTGRSIVSCWPLYMFPQKPNYTEIRQYCNHWRNFADIDDSWKS 240
Qy 241 SILDWTSFNQRIVDVAGPGWNPDMVLVGNFGLSNQOVTOMALWAIMAALPFW 300
Db 241 SILDWTSFNQRIVDVAGPGWNPDMVLVGNFGLSNQOVTOMALWAIMAALPFW 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKGYQLRQGDNFVWVERPLSGLAWAVAMINR 360
Db 301 RHISPOKALLQDKVIAINQDPLGKGYQLRQGDNFVWVERPLSGLAWAVAMINR 360
Qy 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRLGFEYWTSLRSLHNPITGT 417
Db 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRLGFEYWTSLRSLHNPITGT 417
```

Db 61 DSCISEKLFMEAEMLVSEGWKADGAYEYLCDICDWCWAPQDSEGRQADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGKIYADVGNKTCAGFPGSGYDYDIDQTFADGWGVDLLKFGDCCYCDSENL 180
Db 121 ANYVHSGKLGKIYADVGNKTCAGFPGSGYDYDIDQTFADGWGVDLLKFGDCCYCDSENL 180
Qy 181 ADGKMSLALNRGRSIVVSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKMSLALNRGRSIVVSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSFQERIVDVAGPGGNDPDMVLVGNFGLSNQOVTQWALWAIWAAPLFMSNDL 300
Db 241 SILDWTSFQERIVDVAGPGGNDPDMVLVGNFGLSNQOVTQWALWAIWAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQYQLOGDNFEVWERPLSGLAWAVAMINROEIG 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQYQLOGDNFEVWERPLSGLAWAVAMINROEIG 360
Qy 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL 417
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL 417

RESULT 5

US-09-993-059-18
; Sequence 18, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-18

Query Match 100.0%; Score 2293; DB 2; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MOLRNPDLHLCALALRFLALVSDIPGARALDNGLARTPTMGLHWRFCMLDCQEEP 60
Db 1 MOLRNPDLHLCALALRFLALVSDIPGARALDNGLARTPTMGLHWRFCMLDCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKADGAYEYLCDICDWCWAPQDSEGRQADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADGAYEYLCDICDWCWAPQDSEGRQADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGKIYADVGNKTCAGFPGSGYDYDIDQTFADGWGVDLLKFGDCCYCDSENL 180
Db 121 ANYVHSGKLGKIYADVGNKTCAGFPGSGYDYDIDQTFADGWGVDLLKFGDCCYCDSENL 180
Qy 181 ADGKMSLALNRGRSIVVSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKMSLALNRGRSIVVSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSFQERIVDVAGPGGNDPDMVLVGNFGLSNQOVTQWALWAIWAAPLFMSNDL 300
Db 241 SILDWTSFQERIVDVAGPGGNDPDMVLVGNFGLSNQOVTQWALWAIWAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQYQLOGDNFEVWERPLSGLAWAVAMINROEIG 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQYQLOGDNFEVWERPLSGLAWAVAMINROEIG 360

Qy 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL 417
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL 417
RESULT 6
US-10-103-327-18
; Sequence 18, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-18

Query Match 100.0%; Score 2293; DB 2; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MOLRNPDLHLCALALRFLALVSDIPGARALDNGLARTPTMGLHWRFCMLDCQEEP 60
Db 1 MOLRNPDLHLCALALRFLALVSDIPGARALDNGLARTPTMGLHWRFCMLDCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKADGAYEYLCDICDWCWAPQDSEGRQADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADGAYEYLCDICDWCWAPQDSEGRQADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGKIYADVGNKTCAGFPGSGYDYDIDQTFADGWGVDLLKFGDCCYCDSENL 180
Db 121 ANYVHSGKLGKIYADVGNKTCAGFPGSGYDYDIDQTFADGWGVDLLKFGDCCYCDSENL 180
Qy 181 ADGKMSLALNRGRSIVVSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKMSLALNRGRSIVVSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSFQERIVDVAGPGGNDPDMVLVGNFGLSNQOVTQWALWAIWAAPLFMSNDL 300
Db 241 SILDWTSFQERIVDVAGPGGNDPDMVLVGNFGLSNQOVTQWALWAIWAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQYQLOGDNFEVWERPLSGLAWAVAMINROEIG 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQYQLOGDNFEVWERPLSGLAWAVAMINROEIG 360
Qy 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL 417
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL 417

RESULT 7

US-09-993-059-14
; Sequence 14, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPS06
; CURRENT APPLICATION NUMBER: US/09/993,059

```
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-14

Query Match      100.0%; Score 2293; DB 2; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRPFCMLDCQEEP 60
Db 1 MQLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRPFCMLDCQEEP 60

QY 61 DSCISEKLFMEAEMLVSEGWKDGAYEYLICIDCWMAPOQDSGRQLQADPQRPFGIROL 120
Db 61 DSCISEKLFMEAEMLVSEGWKDGAYEYLICIDCWMAPOQDSGRQLQADPQRPFGIROL 120

QY 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGYDYIDAQTFADWGVDDLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGYDYIDAQTFADWGVDDLKFDGCGYCDLSLENL 180

QY 181 ADGYKMSLALNRTGRSIVYSCWPLYMPPFQKPNYTEIROYCNHWRNFADIDDSWKSIX 240
Db 181 ADGYKMSLALNRTGRSIVYSCWPLYMPPFQKPNYTEIROYCNHWRNFADIDDSWKSIX 240

QY 241 SILDWTSFNQERIVDVAGPGGNDPDMVLVGNFGLSNQOQVTOMALWAIWAAPLFWMSNDL 300
Db 241 SILDWTSFNQERIVDVAGPGGNDPDMVLVGNFGLSNQOQVTOMALWAIWAAPLFWMSNDL 300

QY 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRGQDNFVWERPLSLGLAWAVAMINRQEI 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRGQDNFVWERPLSLGLAWAVAMINRQEI 360

QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 417
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 417

RESULT 8
US-10-103-327-14
; Sequence 14, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-14

Query Match      100.0%; Score 2293; DB 2; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRPFCMLDCQEEP 60
Db 1 MQLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRPFCMLDCQEEP 60

QY 61 DSCISEKLFMEAEMLVSEGWKDGAYEYLICIDCWMAPOQDSGRQLQADPQRPFGIROL 120
Db 61 DSCISEKLFMEAEMLVSEGWKDGAYEYLICIDCWMAPOQDSGRQLQADPQRPFGIROL 120

QY 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGYDYIDAQTFADWGVDDLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGYDYIDAQTFADWGVDDLKFDGCGYCDLSLENL 180

QY 181 ADGYKMSLALNRTGRSIVYSCWPLYMPPFQKPNYTEIROYCNHWRNFADIDDSWKSIX 240
Db 181 ADGYKMSLALNRTGRSIVYSCWPLYMPPFQKPNYTEIROYCNHWRNFADIDDSWKSIX 240

QY 241 SILDWTSFNQERIVDVAGPGGNDPDMVLVGNFGLSNQOQVTOMALWAIWAAPLFWMSNDL 300
Db 241 SILDWTSFNQERIVDVAGPGGNDPDMVLVGNFGLSNQOQVTOMALWAIWAAPLFWMSNDL 300

QY 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRGQDNFVWERPLSLGLAWAVAMINRQEI 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRGQDNFVWERPLSLGLAWAVAMINRQEI 360

QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 417
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 417

RESULT 9
US-07-602-824A-2
; Sequence 2, Application US/07602824A
; Patent No. 5356804
; GENERAL INFORMATION:
; APPLICANT: Deanick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/602,824A
; FILING DATE: 24-OCT-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-602-824A-2

Query Match      100.0%; Score 2293; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPHELGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRPFCMLDCQEEP 60
```

Db 1 MOLNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCEEP 60
Qy 61 DSCISEKLFWMAELMVSCEWKDAGYEYLCIDDCWMAPOQDSEGRLOADPQRPFGIRQL 120
Db 61 DSCISEKLFWMAELMVSCEWKDAGYEYLCIDDCWMAPOQDSEGRLOADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGIYADVGNKTCAGPGSGFYDIDAQTFADGWVDLLKFGDCYCDSENL 180
Db 121 ANYVHSGKLGIYADVGNKTCAGPGSGFYDIDAQTFADGWVDLLKFGDCYCDSENL 180
Qy 181 ADGKMSLALNRGRSIVYSCWPLMYPKPNYTEIRQYCNHWNFNADIDDSWSIK 240
Db 181 ADGKMSLALNRGRSIVYSCWPLMYPKPNYTEIRQYCNHWNFNADIDDSWSIK 240
Qy 241 SILDTSFNQERIVDVAGPGGNDPDMLVGNFGLSNQVQTMALWAIMAAPLFMSNDL 300
Db 241 SILDTSFNQERIVDVAGPGGNDPDMLVGNFGLSNQVQTMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKDVIAINQDPLGKQGYLROGDNFEVWERPLSGLAWAVAMINRQIG 360
Db 301 RHISPOKALLQDKDVIAINQDPLGKQGYLROGDNFEVWERPLSGLAWAVAMINRQIG 360
Qy 361 GPRSYYTAVASLGKGVACNACFTITOLLPVKRLGIFYEWSRLRSHINPTGTLLQL 417
Db 361 GPRSYYTAVASLGKGVACNACFTITOLLPVKRLGIFYEWSRLRSHINPTGTLLQL 417

RESULT 10

US-07-602-608-3
; Sequence 3, Application US/07602608
; Patent No. 5382524
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; APPLICANT: Wang, Anne M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY
; TITLE OF INVENTION: ACTIVE ALPHA-N-ACETYL GALACTOSAMINIDASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/07/602,608
; FILING DATE: 24-OCT-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-008
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-07-602-608-3

Query Match 100.0%; Score 2293; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MOLNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCEEP 60
Db 1 MOLNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCEEP 60
Qy 61 DSCISEKLFWMAELMVSCEWKDAGYEYLCIDDCWMAPOQDSEGRLOADPQRPFGIRQL 120
Db 61 DSCISEKLFWMAELMVSCEWKDAGYEYLCIDDCWMAPOQDSEGRLOADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGIYADVGNKTCAGPGSGFYDIDAQTFADGWVDLLKFGDCYCDSENL 180
Db 121 ANYVHSGKLGIYADVGNKTCAGPGSGFYDIDAQTFADGWVDLLKFGDCYCDSENL 180
Qy 181 ADGKMSLALNRGRSIVYSCWPLMYPKPNYTEIRQYCNHWNFNADIDDSWSIK 240
Db 181 ADGKMSLALNRGRSIVYSCWPLMYPKPNYTEIRQYCNHWNFNADIDDSWSIK 240
Qy 241 SILDTSFNQERIVDVAGPGGNDPDMLVGNFGLSNQVQTMALWAIMAAPLFMSNDL 300
Db 241 SILDTSFNQERIVDVAGPGGNDPDMLVGNFGLSNQVQTMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKDVIAINQDPLGKQGYLROGDNFEVWERPLSGLAWAVAMINRQIG 360
Db 301 RHISPOKALLQDKDVIAINQDPLGKQGYLROGDNFEVWERPLSGLAWAVAMINRQIG 360
Qy 361 GPRSYYTAVASLGKGVACNACFTITOLLPVKRLGIFYEWSRLRSHINPTGTLLQL 417
Db 361 GPRSYYTAVASLGKGVACNACFTITOLLPVKRLGIFYEWSRLRSHINPTGTLLQL 417

RESULT 11

US-07-983-451-2
; Sequence 2, Application US/07983451
; Patent No. 5401650
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; TITLE OF INVENTION: Cloning and Expression of Biologically
; TITLE OF INVENTION: Active alpha-Galactosidase A
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/983,451
; FILING DATE: 30-NOV-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 07/983,451
; REFERENCE/DOCKET NUMBER: 6923-030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```
; MOLECULE TYPE: protein
; US-07-983-451-2
;
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-261-578-3

Query Match 100.0%; Score 2293; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPELHLCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCMNCDCQEEP 60
DB 1 MQLRNPELHLCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCMNCDCQEEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPOQDSEGRLOADQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPOQDSEGRLOADQRFPHGIRQL 120
QY 121 ANYVHSGKLGKIYADVGNKTCAGFPGSFGYYIDIAQTADWGVDDLKFGDGCYCDLSLENL 180
DB 121 ANYVHSGKLGKIYADVGNKTCAGFPGSFGYYIDIAQTADWGVDDLKFGDGCYCDLSLENL 180
QY 181 ADGCKHMSLALNRTGRSIVVSCWPLVMPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
DB 181 ADGCKHMSLALNRTGRSIVVSCWPLVMPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
QY 241 SILDWTFSNQRERIVDVAGPGGWNDPDMVLVGNFGLSNQVQTQMALWAIWAAFLFMSNDL 300
DB 241 SILDWTFSNQRERIVDVAGPGGWNDPDMVLVGNFGLSNQVQTQMALWAIWAAFLFMSNDL 300
QY 301 RHISPOKALLQDKVDIATINQDPLGKQGYQLRGDNFEVWERPLSGLAWAVAMINRQIEIG 360
DB 301 RHISPOKALLQDKVDIATINQDPLGKQGYQLRGDNFEVWERPLSGLAWAVAMINRQIEIG 360
QY 361 GPRSYYTIVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLQL 417
DB 361 GPRSYYTIVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLQL 417

RESULT 12
US-08-261-578-3
; Sequence 3, Application US/08261578
; Patent No. 5491075
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; APPLICANT: Wang, Anne M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY
; TITLE OF INVENTION: ACTIVE ALPHA-N-ACETYL GALACTOSAMINIDASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,578
; FILING DATE: 17-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/602,608
; FILING DATE: 24-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
```

```
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-261-578-3

Query Match 100.0%; Score 2293; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPELHLCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCMNCDCQEEP 60
DB 1 MQLRNPELHLCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCMNCDCQEEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPOQDSEGRLOADQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPOQDSEGRLOADQRFPHGIRQL 120
QY 121 ANYVHSGKLGKIYADVGNKTCAGFPGSFGYYIDIAQTADWGVDDLKFGDGCYCDLSLENL 180
DB 121 ANYVHSGKLGKIYADVGNKTCAGFPGSFGYYIDIAQTADWGVDDLKFGDGCYCDLSLENL 180
QY 181 ADGCKHMSLALNRTGRSIVVSCWPLVMPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
DB 181 ADGCKHMSLALNRTGRSIVVSCWPLVMPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
QY 241 SILDWTFSNQRERIVDVAGPGGWNDPDMVLVGNFGLSNQVQTQMALWAIWAAFLFMSNDL 300
DB 241 SILDWTFSNQRERIVDVAGPGGWNDPDMVLVGNFGLSNQVQTQMALWAIWAAFLFMSNDL 300
QY 301 RHISPOKALLQDKVDIATINQDPLGKQGYQLRGDNFEVWERPLSGLAWAVAMINRQIEIG 360
DB 301 RHISPOKALLQDKVDIATINQDPLGKQGYQLRGDNFEVWERPLSGLAWAVAMINRQIEIG 360
QY 361 GPRSYYTIVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLQL 417
DB 361 GPRSYYTIVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLQL 417

RESULT 13
US-08-261-577-7
; Sequence 7, Application US/08261577
; Patent No. 5580757
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; TITLE OF INVENTION: Cloning and Expression of Biologically
; TITLE OF INVENTION: Active alpha-Galactosidase A
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,577
; FILING DATE: 17-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
```

```
;
;
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-042
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-261-577-7

Query Match 100.0%; Score 2293; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPGLHGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Db 1 MQLRNPGLHGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKDAYEYLCTDCCWMAPOQDSGRLOADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKDAYEYLCTDCCWMAPOQDSGRLOADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGLIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFGDCYCDLSLENL 180
Db 121 ANYVHSGKLGLIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFGDCYCDLSLENL 180
Qy 181 ADGKMSLALNRTGRSIVSCWPLYMPPQKPNYTEIRQYCNHWNFNADIDDSWKSIIK 240
Db 181 ADGKMSLALNRTGRSIVSCWPLYMPPQKPNYTEIRQYCNHWNFNADIDDSWKSIIK 240
Qy 241 SILDWTFSFQNERIVDVAGPGGWNDDPMLVIGNFGLSWQVQTMALWAIAPLFSNNDL 300
Db 241 SILDWTFSFQNERIVDVAGPGGWNDDPMLVIGNFGLSWQVQTMALWAIAPLFSNNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQGDNEFVWERPLSLGLAWAVAMINRQEIG 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQGDNEFVWERPLSLGLAWAVAMINRQEIG 360
Qy 361 GPRSYTIAVASLGKGVACNPFACITQLLPVKRKLGFYEWTSRLRSHINPTGTLLQL 417
Db 361 GPRSYTIAVASLGKGVACNPFACITQLLPVKRKLGFYEWTSRLRSHINPTGTLLQL 417

RESULT 14
US-08-261-577-9
; Sequence 9, Application US/08261577
; Patent No. 5580757
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David P.
; APPLICANT: Ioannou, Yiannis A.
; TITLE OF INVENTION: Cloning and Expression of Biologically
; TITLE OF INVENTION: Active alpha-Galactosidase A
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,577
; FILING DATE: 17-JUN-1994

;
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-042
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-261-577-9

Query Match 100.0%; Score 2293; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPGLHGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Db 1 MQLRNPGLHGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKDAYEYLCTDCCWMAPOQDSGRLOADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKDAYEYLCTDCCWMAPOQDSGRLOADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGLIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFGDCYCDLSLENL 180
Db 121 ANYVHSGKLGLIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFGDCYCDLSLENL 180
Qy 181 ADGKMSLALNRTGRSIVSCWPLYMPPQKPNYTEIRQYCNHWNFNADIDDSWKSIIK 240
Db 181 ADGKMSLALNRTGRSIVSCWPLYMPPQKPNYTEIRQYCNHWNFNADIDDSWKSIIK 240
Qy 241 SILDWTFSFQNERIVDVAGPGGWNDDPMLVIGNFGLSWQVQTMALWAIAPLFSNNDL 300
Db 241 SILDWTFSFQNERIVDVAGPGGWNDDPMLVIGNFGLSWQVQTMALWAIAPLFSNNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQGDNEFVWERPLSLGLAWAVAMINRQEIG 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQGDNEFVWERPLSLGLAWAVAMINRQEIG 360
Qy 361 GPRSYTIAVASLGKGVACNPFACITQLLPVKRKLGFYEWTSRLRSHINPTGTLLQL 417
Db 361 GPRSYTIAVASLGKGVACNPFACITQLLPVKRKLGFYEWTSRLRSHINPTGTLLQL 417

RESULT 15
US-09-993-059-10
; Sequence 10, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-993-059-10

Query Match 100.0%; Score 2293; DB 2; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
```

Matches 417;		Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	MQLRNPEHLHGCALALRFLALVSWDIFGARALDNGLARTPTMGWLHWERFMCNLDCEEP 60							
Db	1								
Qy	61	DSCISEKLFMEMAELMVSEGWKDGAYEYLCIDDCWMAQORDSEGRLOADPQRPFGIRQL 120							
Db	61								
Qy	121	ANYVHSKGLKGIYADYGNKTCAGFPGSGFYDIDAQTFADWGVDDLKFDGCYCDLSLENL 180							
Db	121								
Qy	181	ADGKHMSLALNRTGRSIVYSCWPLYMPPQKPNYTEIRQYCNHWNFPADIDDSWKSIIK 240							
Db	181								
Qy	241	SILDWTSFNQERIVDVAGPGGNDPDMLVIGNFGLSNQVOTQWALWAIMAAPLFMSNDL 300							
Db	241								
Qy	301	RHISPOAKALLQDKDVIATINQDPLGKQGYQLRQGDNFEVWERPLSGLAWAVAMINRQEIIG 360							
Db	301								
Qy	361	GPRSYYTIAVASLKGAVACNPACFITOLLPVKRKLGIFYEWTSLRSHINPTGTVLLQL 417							
Db	361								

Search completed: December 31, 2005, 23:48:33
Job time : 45.1379 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 23:16:51 ; Search time 33.5517 Seconds
(without alignments)
1195.837 Million cell updates/sec

Title: US-10-602-220-16
Perfect score: 2293
Sequence: 1 MQLRNPGLHLCALRFLA.....EWSRLRSHNP7GTLLQL 417

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_80.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2293	100.0	429	1 GBHUA	alpha-galactosidas
2	1806	78.8	419	2 JC4532	alpha-galactosidas
3	1051	45.8	405	2 S45522	alpha-N-acetylglala
4	952.5	41.5	411	2 A35485	alpha-N-acetylglala
5	930	40.6	358	2 A33265	alpha-N-acetylglala
6	861	37.5	451	2 T24018	hypothetical prote
7	716.5	31.2	434	2 T47748	alpha-galactosidas
8	712	31.1	378	2 T50781	alpha-galactosidas
9	687.5	30.0	422	2 T06388	alpha-galactosidas
10	682	29.7	425	2 T10850	alpha-galactosidas
11	681.5	29.7	411	2 S07472	alpha-galactosidas
12	676	29.5	680	2 T36472	probable secreted
13	609	26.6	436	2 T39118	probable alpha-gal
14	598	26.1	396	2 JC5558	alpha-galactosidas
15	557	24.3	469	2 S45453	alpha-galactosidas
16	543.5	23.7	545	2 S23582	alpha-galactosidas
17	521	22.7	471	2 S50312	alpha-galactosidas
18	519	22.6	471	2 S50312	alpha-galactosidas
19	516	22.5	444	2 S74221	alpha-galactosidas
20	514	22.4	471	2 TQ1021	alpha-galactosidas
21	510	22.2	471	2 S50310	alpha-galactosidas
22	502	21.9	471	1 GBBYAG	probable alpha-gal
23	300.5	13.1	204	2 T04423	hypothetical prote
24	282.5	12.3	432	2 F83883	hypothetical prote
25	198	8.6	159	2 T04422	alpha-galactosidas
26	180.5	7.9	624	2 S74222	alpha-galactosidas
27	129	5.6	641	2 A55549	glucan 1,6-alpha-l
28	125	5.5	348	2 A70311	hypothetical prote
29	116	5.1	4199	2 S76412	hypothetical prote

RESULT 1

GBHUA

alpha-galactosidase (EC 3.2.1.22) A precursor - human
N/Alternate names: alpha-D-galactoside galactohydrolase; melibiase
C/Species: Homo sapiens (man)
C/Date: 30-Jun-1987 #sequence revision 27-Oct-1995 #text change 09-Jul-2004
C/Accession: S04081; A29608; A30214; S14879; A00896; B00896; I37140
R/Kornreich, R.; Desnick, R.J.; Bishop, D.F.

Nucleic Acids Res. 17, 3301-3302, 1989

A/Title: Nucleotide sequence of the human alpha-galactosidase A gene.

A/Reference number: S04081; MUID:89263745; PMID:2542896

A/Accession: S04081

A/Status: translation not shown

A/Molecule type: DNA

A/Residues: 1-429 <KOR>

A/Cross-references: UNIPROT:P06280; UNIPARC:UPI0000033A30; EMBL:X14448; NID:g31755; PIDN:

R/Quinn, M.; Hantzopoulos, P.; Fidanza, V.; Calhoun, D.H.

Gene 58, 177-188, 1987

A/Title: A genomic clone containing the promoter for the gene encoding the human lysosom

A/Reference number: A29608; MUID:88112869; PMID:2892762

A/Accession: A29608

A/Molecule type: DNA

A/Residues: 1-64 <QUI>

A/Cross-references: UNIPARC:UPI000016A96A; GB:M18242; NID:g182944; PIDN:AAA52514.1; PID:

R/Bishop, D.F.; Kornreich, R.; Desnick, R.J.

Proc. Natl. Acad. Sci. U.S.A. 85, 3903-3907, 1988

A/Title: Structural organization of the human alpha-galactosidase A gene: further eviden

A/Reference number: A30214; MUID:88234528; PMID:2836863

A/Accession: A30214

A/Molecule type: DNA

A/Residues: 1-64 <BIS>

A/Cross-references: UNIPARC:UPI000016A96A; EMBL:M20317; EMBL:J03249

R/Koide, T.; Ishiura, M.; Iwai, K.; Inoue, M.; Kaneda, Y.; Okada, Y.; Uchida, T.

FEBS Lett. 259, 353-356, 1990

A/Title: A case of Fabry's disease in a patient with no alpha-galactosidase A activity c

A/Reference number: S14879; MUID:90092580; PMID:2152885

A/Accession: S14879

A/Molecule type: mRNA

A/Residues: 1, K, 3-39, 'S', 41-429 <KOI>

A/Cross-references: UNIPARC:UPI000011E285; EMBL:X16889

A/Experimental source: Fabry's disease patient

R/Bishop, D.F.; Calhoun, D.H.; Bernstein, H.S.; Hantzopoulos, P.; Quinn, M.; Desnick, R.

Proc. Natl. Acad. Sci. U.S.A. 83, 4859-4863, 1986

A/Title: Human alpha-galactosidase A: nucleotide sequence of a cDNA clone encoding the m

A/Reference number: A00896; MUID:86259694; PMID:3014515

A/Accession: A00896

A/Molecule type: mRNA

A/Residues: 27-429 <BI2>

A/Cross-references: UNIPARC:UPI0000000358; GB:M13571; NID:g178245; PIDN:AAA51676.1; PID:

A/Experimental source: lung

A/Accession: B00896

A/Molecule type: protein

C;Genetics:
A;Gene: GDB:NAGA
A;Cross-references: GDB:119445; OMIM:104170
A;Map position: 22q11-22q11
A;Introns: 6/1; 51/2; 108/3
C;Superfamily: alpha-galactosidase
C;Keywords: alternative splicing; glycoprotein; glycosidase; hydrolase; lysosome
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-41/Product: alpha-N-acetylgalactosaminidase, splice form 2 #status predicted <MAT>
F;124,177,201,359,385,391/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 41.5%; Score 952.5; DB 2; Length 411;
Best Local Similarity 50.0%; Pred. No. 6.7e-74;
Matches 198; Conservative 55; Mismatches 120; Indels 23; Gaps 9;

Qy 32 LDNGLARTPTTGWLHWFRCNLDQCEPDPSCISSEKLFMEMAELMVSEGWKDAYEYLICI 91
Dy 18 LDNGLQTTPNGMLAWERFCNINCEDEPKNCISEQLFMEMADRMADQCHRDNGYTYLNI 77
Qy 92 DDCWMAQRDSEGRQLADPQRPPIHGIROLQANTVHSKGLGIYADVGNKTCAGPPG-SFG 150
Dy 78 DDCWIG-GRDASGRMLPDKRPPIHGIPLADYVLSGLGIYADMGNFCTMGYPGTLD 136
Qy 151 YYDTDACTFADMGVLLKFDGDCYCDSENLADGVKXMSALNRTGSRIVYVSCWELPMYP 210
Dy 137 KWQDAQTFAEWKVDMLKLDGCF-STPERAAGYPRMAALNATGPASFCSWPAYEGG 195
Qy 211 F-QKPNYTEIRQYCNHNRNFADIDDSWSKSIKLDWTSFNQERIVDVAGPGWNDPMLV 269
Dy 196 LPPRVNYSLLADICNLWRNYDDIQDSWSVLSILNFWVEHQDILQPVAGPGHNDPMLL 255
Qy 270 IGNFGLSNQOVTQWALMAIAPLMSNDLRHSIPQAKALLQDKVIAINQDPLGKQGY 329
Dy 256 IGNFGLSLEQSRQWALMTVLAAPLLMSDTRTISAQNMDDILQNPLMIKINQDPLGIQGR 315
Qy 330 QL-RQGNFVEFWERPLSLAVAMIN-ROEIGGPRSYTIAVASLKGVCNCPACFITQL 387
Dy 316 RIHKESLIEVYMPRLSKASALVFFSCRTDM-PYRHSSILQLN-----FTGSV 364
Qy 388 LPVKKRLGFBWTSKRLSH-----INPTGTVLQL 417
Dy 365 IYEADQVYSGDIISGLRDETFTVINSGVVMWYL 400

RESULT 5
A33265
alpha-N-acetylgalactosaminidase (EC 3.2.1.49) precursor, splice form 1 - human
N;Alternate names: alpha-galactosidase B
C;Species: Homo sapiens (man)
C;Date: 21-Feb-1990 #sequence_revision 21-Feb-1990 #text_change 09-Jul-2004
C;Accession: A33265
R;Tsuiji, S.; Yamauchi, T.; Hiraiwa, M.; Isobe, T.; Okuyama, T.; Sakimura, K.; Takahashi, K.;
Biochem. Biophys. Res. Commun. 163, 1498-1504, 1989
A;Title: Molecular cloning of a full-length cDNA for human alpha-N-acetylgalactosaminidase
A;Reference number: A33265; MUID:89392067; PMID:2551294
A;Accession: A33265
A;Molecule type: mRNA
A;Residues: 1-358 <TSU>
A;Cross-references: UNIPROT:P17050; UNIPARC:UPI000016ADB4; GB:M29276; NID:gl89052; PIDN:
A;Experimental source: clone pcd-HS1204
C;Genetics:
A;Gene: GDB:NAGA
A;Cross-references: GDB:119445; OMIM:104170
A;Map position: 22q11-22q11
A;Introns: 6/1; 51/2; 108/3; 168/1; 199/3; 253/3; 319/3
C;Superfamily: alpha-galactosidase
C;Keywords: alternative splicing; glycoprotein; glycosidase; hydrolase; lysosome
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-358/Product: alpha-N-acetylgalactosaminidase, splice form 1 #status predicted <MAT>
F;124,177,201/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 40.6%; Score 930; DB 2; Length 358;
Best Local Similarity 57.7%; Pred. No. 4.8e-72;
Matches 198; Conservative 55; Mismatches 120; Indels 23; Gaps 9;

	Matches	176;	Conservative	40;	Mismatches	85;	Indels	4;	Gaps	4;
Qy	32	LDNGLARTPTMGTLHWRFMCNLDCEBPDS	CISEKLFMEAEIWMSEGMKADGYELCI	91						
Db	18	LDNGLLTQPPMGWLAWFRFCNINCEDEDPKNCISEQLFMEWADRMADQGWDRMGTYILNI	77							
Qy	92	DDCWMAFORDBEGRLOADPQRPFHIGIRQLANYVHSGKLGIYADVGNKTCAGPFG-SFG	150							
Db	78	DDCWIG-GRDASGRMLPDPKRPFGHIFPLADYVHSLGLKGIYADMGNFTCMYPGTITLD	136							
Qy	151	YYDIDAQTADPAGVGYDLKPFDCYCDSELENLADGKYKMSLALNRTGRSIVSYSCWPLMYMP	210							
Db	137	KVQDQATQTEAFWKVMDLKLDCGF-STPEEAQGYGPKMAAGLNATGRPIAFSCSWPAYEGG	195							
Qy	211	P-QKPNYTEIRQYCNHNRNFADIDDKSKSIILDWTISFNQERTVDVAGPGGWNDDPMLV	269							
Db	196	LPFRVNTYSLQADI CNLWRNYDDIQDSWRKSVLJLNWFVEHQDIIQPVAGPGHWNDDPMLL	255							
Qy	270	IGNFGLSWNQVTOQWALWAIMAAPLFMSNDLHRISPOKALLQDKVIAINQDPLGKQGY	329							
Db	256	IGNFGLSELSQSRQAWLWTVLAPLAAFMSTMDTLRTISAQNMMDILQNPMLTKINQDPLGIOGR	315							
Qy	330	QLRQG 334								
Db	316	RIHKG 320								

RESULT 6
T24018 hypothetical protein R07B7.11 - *Caenorhabditis elegans*
C/Species: *Caenorhabditis elegans*
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C/Accession: T24018
R/Harris, B.
submitted to the EMBL Data Library, July 1996
A/Reference number: Z19830
A/Accession: T24018
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-451 <WTL>
A/Cross-references: UNIPROT:Q21801; UNIPARC:UP1000007F870; EMBL:Z75955; PIDN:CAB00120.1
A/Experimental source: clone R07B7
C/Genetics:
A/Gene: CESP:R07B7.11
A/Map position: 5
A/Introns: 61/1; 107/3; 317/3; 351/3
C/Superfamily: alpha-galactosidase

Dn 309 GIMGRLLNSTDIGVYVKQTIPSGDKKSFAYLNRRKEGKYRIETQLASIG---LTD 365
| | : :: : | : | : | : || :

Qy 380 PACFITQLLPVRKLGFYEWTSLRSHINPTGTVLQG 416
|| : : | : | : | :

Db 366 PAGYYVDHINSVLDGLLRSGDSIWSTAPAGSVFFR 402
| : : | : | : | :

RESULT 7

T47748
alpha-galactosidase-like protein - Arabidopsis thaliana
N;Alternate names: protein F18021.270
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C;Accession: T47748
E;Benes, V.; Wurmbach, E.; Drzonek, H.; Ansoerge, W.; Mewes, H.W.; Rudd, S.; Leberer,
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24474
A;Accession: T47748
A;Status: preliminary
A:Molecule type: DNA
A;Residues: 1-434 <BN>
A;Cross-references: UNIPROT-Q9LYL2; UNIPARC:UIP00000A62BB; EMBL:ATF18021
A;Experimental source: cultivar Columbia; BAC clone F18021
C;Genetics:
A;Map position: 3
A;Introns: 73/2; 93/1; 112/1; 123/3; 158/2; 182/3; 202/2; 224/2; 252/2; 274/1;
A;Note: F18021.270
C;Superfamily: alpha-galactosidase

Query Match 31.2%; Score 716.5; DB 2; Length 434;
Best Local Similarity 42.4%; Pred. No. 1.3e-53;
Matches 153; Conservative 54; Mismatches 107; Indels 47; Gaps
:

```

RESULT 8
T50781
alpha-galactosidase (EC 3.2.1.22) [imported] - coffee
C:Species: Coffea arabica (coffee)
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004
C:Accession: T50781
R:Zhu, A.; Goldstein, J.
Gene 140, 227-231, 1994
A:Title: Cloning and functional expression of a cDNA encoding coffee bean
A:Reference number: Z25235; MUID:94193002; PMID:8144030

```

A:Accession: T50781
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-378 <ZHU>
A:Cross-references: UNIPROT:Q42656; UNIPARC:UPI00001256A8; EMBL:L27992; PIDN:AAA33022.1
C:Superfamily: alpha-galactosidase
C:Keywords: glycosidase; hydrolase

Query Match 31.1%; Score 712; DB 2; Length 378;
Best Local Similarity 42.5%; Pred. No. 2.7e-53;
Matches 152; Conservative 51; Mismatches 113; Indels 42; Gaps 8;

QY 32 LDNGLARTPTMGWLHWFRCNLDQCBEPCDSCISEKLFMEAEMLVSEGWKDGAYEYLICI 91
DB 16 LANGLGLTPPMGWSNNHFRCLND-----EKLLRETADAMVSKGLAALGYKYINL 65

QY 92 DDCWMAPODSEGRLOADQPPPHGIROLANYVHSGKLGKIYADVGNKTCAGPFGSFG 150
DB 66 DDCWALNRDSQGNLVKPSGIFPSGIGKALADYVHSGKLGKIYSDAGTQTCSTKMPGSLG 125

QY 151 YYDIDAQTFADGVDLKFDGVCYCDLSLENLADGYKHMSLALNRTGRSIVYS-CEWPLYMW 209
DB 156 HEEQDAKTFASGVVDYLYKDYCNNNNI-SPKERYPIMSKALLNSGRSIFFSLCBWG---- 180

QY 210 PFQKPNYTEIRQYCNHWRNFADIDDSWKSISKILD-----WTSFQNERIVDVAGPGGWNNDP 265
DB 181 ---BEDPATWAEVGVNSWRTTGDIDDSWSGMSRTSRADMDNDKWASY-----AGPGGWNNDP 230

QY 266 DMLVGNFGLSNQVQVQTMALWAIWAAPLFWNSDLRHISPOKALLQDKDVIAINODPLG 325
DB 231 DMLVEVNGGMMTTEYRSHFSIWAALAKAPLLIGCDIRSDMGATFQLLSNAEVIANVQDKLUG 290

QY 326 KQGYQLROQDNFVEWPERPLSGLAWAVAMINR-----QEIGGPRSYTIAVASL 372
DB 291 VQGNKVYKTYGDLVWAGPLSGKRAVALWNRGSSSTATITAYWSDVGLSTAVVNRDL 348

RESULT 9
T06388
alpha-galactosidase (EC 3.2.1.22) - soybean
C:Species: Glycine max (soybean)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 09-Jul-2004
C:Accession: T06388
R:Davis, M.O.; Walker, J.C.; Smith, D.
A:Description: Cloning and expression of a soybean alpha galactosidase gene.
A:Reference number: Z15645
A:Accession: T06388
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-422 <DAV>
A:Cross-references: UNIPROT:Q39811; UNIPARC:UPI00000A7200; EMBL:U12926; PID:G927574; PID:G927574;
A:Experimental source: strain williams
C:Function:
A:Description: catalyzes hydrolysis of melibiose into galactose and glucose
C:Superfamily: alpha-galactosidase
C:Keywords: glycosidase; hydrolase

Query Match 30.0%; Score 687.5; DB 2; Length 422;
Best Local Similarity 43.2%; Pred. No. 4e-51;
Matches 143; Conservative 45; Mismatches 112; Indels 31; Gaps 7;

QY 32 LDNGLARTPTMGWLHWFRCNLDQCBEPCDSCISEKLFMEAEMLVSEGWKDGAYEYLICI 91
DB 60 LDNGLGHTPPMGWSNNHFCN-----IKEDLLRETADAMVSTGLAALGYQYINI 109

QY 92 DDCWMAPODSEGRLOADQPPPHGIROLANYVHSGKLGKIYADVGNKTCAGPFGSFG 150
DB 110 DDCWELNRDSKGNLVKPSGIFPSGIGKALADYVHSGKLGKIYSDAGTQTCSTKMPGSLG 169

QY 151 YYDIDAQTFADGVDLKFDGVCYCDLSLENLADGYKHMSLALNRTGRSIVYS-CEWPLYMW 209
DB 170 HEEQDAKTFASGVVDYLYKDYCNNNNI-SPKERYPIMSEALANTGRPFFSLCBWG---- 224

A:Accession: T50781
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-378 <ZHU>
A:Cross-references: UNIPROT:Q42656; UNIPARC:UPI00001256A8; EMBL:L27992; PIDN:AAA33022.1
C:Superfamily: alpha-galactosidase
C:Keywords: glycosidase; hydrolase

Query Match 31.1%; Score 712; DB 2; Length 378;
Best Local Similarity 42.5%; Pred. No. 2.7e-53;
Matches 152; Conservative 51; Mismatches 113; Indels 42; Gaps 8;

QY 32 LDNGLARTPTMGWLHWFRCNLDQCBEPCDSCISEKLFMEAEMLVSEGWKDGAYEYLICI 91
DB 16 LANGLGLTPPMGWSNNHFRCLND-----EKLLRETADAMVSKGLAALGYKYINL 65

QY 92 DDCWMAPODSEGRLOADQPPPHGIROLANYVHSGKLGKIYADVGNKTCAGPFGSFG 150
DB 66 DDCWALNRDSQGNLVKPSGIFPSGIGKALADYVHSGKLGKIYSDAGTQTCSTKMPGSLG 125

QY 151 YYDIDAQTFADGVDLKFDGVCYCDLSLENLADGYKHMSLALNRTGRSIVYS-CEWPLYMW 209
DB 156 HEEQDAKTFASGVVDYLYKDYCNNNNI-SPKERYPIMSKALLNSGRSIFFSLCBWG---- 180

QY 210 PFQKPNYTEIRQYCNHWRNFADIDDSWKSISKILD-----WTSFQNERIVDVAGPGGWNNDP 265
DB 181 ---BEDPATWAEVGVNSWRTTGDIDDSWSGMSRTSRADMDNDKWASY-----AGPGGWNNDP 230

QY 266 DMLVGNFGLSNQVQVQTMALWAIWAAPLFWNSDLRHISPOKALLQDKDVIAINODPLG 325
DB 231 DMLVEVNGGMMTTEYRSHFSIWAALAKAPLLIGCDIRSDMGATFQLLSNAEVIANVQDKLUG 290

QY 326 KQGYQLROQDNFVEWPERPLSGLAWAVAMINR-----QEIGGPRSYTIAVASL 372
DB 291 VQGNKVYKTYGDLVWAGPLSGKRAVALWNRGSSSTATITAYWSDVGLSTAVVNRDL 348

RESULT 9
T06388
alpha-galactosidase (EC 3.2.1.22) - soybean
C:Species: Glycine max (soybean)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 09-Jul-2004
C:Accession: T06388
R:Davis, M.O.; Walker, J.C.; Smith, D.
A:Description: Cloning and expression of a soybean alpha galactosidase gene.
A:Reference number: Z15645
A:Accession: T06388
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-422 <DAV>
A:Cross-references: UNIPROT:Q39811; UNIPARC:UPI00000A7200; EMBL:U12926; PID:G927574; PID:G927574;
A:Experimental source: strain williams
C:Function:
A:Description: catalyzes hydrolysis of melibiose into galactose and glucose
C:Superfamily: alpha-galactosidase
C:Keywords: glycosidase; hydrolase

Query Match 30.0%; Score 687.5; DB 2; Length 422;
Best Local Similarity 43.2%; Pred. No. 4e-51;
Matches 143; Conservative 45; Mismatches 112; Indels 31; Gaps 7;

QY 32 LDNGLARTPTMGWLHWFRCNLDQCBEPCDSCISEKLFMEAEMLVSEGWKDGAYEYLICI 91
DB 60 LDNGLGHTPPMGWSNNHFCN-----IKEDLLRETADAMVSTGLAALGYQYINI 109

QY 92 DDCWMAPODSEGRLOADQPPPHGIROLANYVHSGKLGKIYADVGNKTCAGPFGSFG 150
DB 110 DDCWELNRDSKGNLVKPSGIFPSGIGKALADYVHSGKLGKIYSDAGTQTCSTKMPGSLG 169

QY 151 YYDIDAQTFADGVDLKFDGVCYCDLSLENLADGYKHMSLALNRTGRSIVYS-CEWPLYMW 209
DB 170 HEEQDAKTFASGVVDYLYKDYCNNNNI-SPKERYPIMSEALANTGRPFFSLCBWG---- 224

QY 210 PFQKPNYTEIRQYCNHWRNFADIDDSWKSISKILD-----WTSFQNERIVDVAGPGGWNNDP 265
DB 225 ---BEDPATWAKSVGNSWRTTGDIDQKWDMSISRADLNDKWASY-----AGPGGWNNDP 274

QY 266 DMLVGNFGLSNQVQVQTMALWAIWAAPLFWNSDLRHISPOKALLQDKDVIAINODPLG 325
DB 275 DMLVEVNGGMMTTEYRSHFSIWAALAKAPLLIGCDIRALDATTKELLSNKEVIANVQDKLUG 334

QY 326 KQGYQLROQDNFVEWPERPLSGLAWAVAMINR 356
DB 335 VQGNKVYKTYGDLVWAGPLSGKRAVALWNR 365

RESULT 10
T10860
alpha-galactosidase (EC 3.2.1.22) - kidney bean
C:Species: Phaseolus vulgaris (kidney bean)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C:Accession: T10860
R:Davis, M.O.; Walker, J.C.; Smith, D.
A:Description: Cloning and expression of a pinto bean alpha galactosidase gene.
A:Reference number: Z17189
A:Accession: T10860
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-425 <DAV>
A:Cross-references: UNIPROT:Q41100; UNIPARC:UPI00000A0F13; EMBL:U12927; NID:G927576; PID:G927576;
C:Function:
A:Description: catalyzes hydrolysis of melibiose into galactose and glucose
C:Superfamily: alpha-galactosidase
C:Keywords: glycosidase; hydrolase

Query Match 29.7%; Score 682; DB 2; Length 425;
Best Local Similarity 41.5%; Pred. No. 1.2e-50;
Matches 144; Conservative 54; Mismatches 113; Indels 36; Gaps 9;

QY 19 LALVSWDIPGARAL-DNGLARTPTMGWLHWFRCNLDQCBEPCDSCISEKLFMEAEMLV 77
DB 49 MMMMSREVDHRRNLVGNGLGQTPPMGWSNNHFCN-----INEDLIRETADAMV 98

QY 78 SEGKWDAGYEYLICDDCWMAPODSEGRLOADQPPPHGIROLANYVHSGKLGKIYADV 137
DB 99 STGLAALGYQYINIDDCWELNRDSQGNLVKPSGIFPSGIGKALADYVHSGKLGKIYSDA 158

QY 138 GNKTCAGPFGSFGYDIDAQTFADGVDLKFDGVCYCDLSLENLA--DGYKHMSLALNRT 194
DB 159 GTQTCSTKMPGSLGHEBQDAKTFASWIDYLYKDYCN-----ENKNISPKERYPPMSKALANS 215

QY 195 GRSIVYS-CEWPLYMWPFQKPNYTEIRQYCNHWRNFADIDDSWKSISKILD-----WTSFN 249
DB 216 GRPIFFSLCBWG-----SEDPATWAKSVGNSWRTTGDIEDKWSMISRADLNDWSASY- 268

QY 250 QERIVDVAGPGGWNNDPMLVITGNFGLSNQVQVQTMALWAIWAAPLFWNSDLRHISPOKA 309
DB 269 -----AGPGGWNNDPDMLEVGNGGMMTTEYRSHFSIWAALAKAPLLIGCDIRALDVTTC 321

QY 310 LLQDKDVIAINODPLGKQGYQLROQDNFVEWPERPLSGLAWAVAMINR 356
DB 322 LLSNEEVIANVQDKLUGVQGNKVYKTYGDLVWAGPLSGKRAVALWNR 368

RESULT 11
S07472
alpha-galactosidase (EC 3.2.1.22) precursor - guar
C:Species: Cyamopsis tetragonoloba (guar, cluster bean)
C:Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 09-Jul-2004
C:Accession: S07472
R:Overbeek, N.; Fellinger, A.J.; Toonen, M.Y.; van Wassenaar, D.; Verrips, C.T.
Plant Mol. Biol. 13, 541-550, 1989
A:Title: Cloning and nucleotide sequence of the alpha-galactosidase cDNA from Cyamopsis
A:Reference number: S07472; MUID:91370836; PMID:2577496

A:Accession: S07472
A:Molecule type: mRNA
A:Residues: 1-411 <OV>
A:Cross-references: UNIPROT:P14749; UNIPARC:UPI00001256A9; EMBL:X14619; NID:g18291; PIDN
C:Note: the authors translated the codon ATT for residue 20 as Asn, TAT for residue 140
C:Superfamily: alpha-galactosidase
C:Keywords: glycosidase; hydrolase

Query Match 29.7%; Score 681.5; DB 2; Length 411;
Best Local Similarity 42.2%; Pred. No. 1.3e-50;
Matches 141; Conservative 52; Mismatches 102; Indels 39; Gaps 9;
QY 33 DNGIARTTGMGLHWRFCMLDQCQEPDSCISEKLFMEAEMLVSEGWKAGYEYLCID 92
DB 49 ENGLGQTPPMGWSNNHFGCD-----INENVRETADAMVSTGLAALGYQYINLD 98
QY 93 DCWMAPOQDSRGLQADPQRPFGHGIROLANYVHSGKLKLGIVADVGNKTCAGPFGSGFY 151
DB 99 DCWAEALNRDSEGNMVPNAAPFSGIKALADYVHSGKLKLGIVSDAGNQTCSCRMPGSLGH 158
QY 152 YDIDAQTFADWGVLLKFDGVCYCDLENLA---DGYKHMSLALNRTGRSIVYS-CEWPL 206
DB 159 EEQDAKTPASGVVLYKYDNC-----ENLGISVVERPPMGKALLSSGRPIFFSNCE--- 210
QY 207 YMWPFQKPNYTEIROCYCNHWNFNADIDDSMKSIKSIILD----WTSFNOERIVDVAGPGGW 262
DB 211 --WGVEDPQIW-AKSGNSWTTGDIEDNWSMTSIADSNKWSY-----AGPGGW 259
QY 263 NDPMLVIGNFGLSNQOVTQMAIWAIPAFLFWSNDRHISPOAKALLOKQVIAINQD 322
DB 260 NDPDMLVGVNGWMTTEYRSHFSIWALAKAPLLVGCDIRAMDDTTHELISNAEVIANNQD 319
QY 323 PLGKQGYLROGDNFEVWERPLSGLAWAVAMINR 356
DB 320 KLGVGKQKVKSTNDLEVWAGPLSDNKVAIVLWNR 353

RESULT 12

T36472
probable secreted alpha-galactosidase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T36472
R:Seeger, K.J.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1999
A:Reference number: Z21607
A:Accession: T36472
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-680 <SE>
A:Cross-references: UNIPROT:Q9S2C9; UNIPARC:UPI00000DB403; EMBL:AL110470; PIDN: CAB54169
A:Experimental source: strain A3(2)
C:Genetics: SCODB:SCF85.12

Query Match 29.5%; Score 676; DB 2; Length 680;
Best Local Similarity 38.2%; Pred. No. 7.3e-50;
Matches 161; Conservative 67; Mismatches 140; Indels 54; Gaps 14;
QY 11 GCALALRFLALVSWDIPGARALDNGIARTPTMGWLHWRFCMLDQCQEPDSCISEKLFM 70
DB 29 GLAVSAQVAPAAAPDPAPASSGHEGLALTTPMGFNWNN---STHCRDE---FNESMVK 80
QY 71 EMAELMVSSEGWKAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGHGIROLANYVHSGKLK 130
DB 81 GIADLFVERGLKAGYEYVNLDDCWALPERDAGRLVDPKRPFGNGIEAVADYVHSGKLK 140
QY 131 LGIYADVGNKTCAGPFGSGFYVDIDAQTFADWGVLLKFDGVCYCDLENLAADGYKMS 188
DB 141 FGIYTSAGTKTCSSIGFPEALGHEYSDAQFADWGVVLYKYDNCNNQGVDAKORYTMR 199
QY 189 LALNRTGRSIVYS-CEW----PLYMWPFOKPNYTEIROCYCNHWNFNADIDDSWKSISIL 243

DB 200 DALAATGRPIVYSICEWGENKP---WEWA---GDLOQL---WRTTGDIINDSWGSMSSIM 249
QY 244 DWTSFNOERIVDVAGPGGWNDPDMVLVIGNFGLSNQOVTQMAIWAIPAFLFWSNDRHIS 303
DB 250 K----SNLEAEYARPGWNDPDMLEVGNGWMTTEYRSHFSMWISIWAAPLLIGTDLRTA 305
QY 304 SPOAKALLOKQVIAINQDPLGKQGYLROGDNFEVWERPLSGLAWAVAMINRQIEGGPR 363
DB 306 PESAFEILTNDVEIVADQDPLGKQGEVVSSEGGRWVSKELADGSRVALFNE-----GSR 361
QY 364 SYTTAVASLKGVCNACPAFITQLLPVKKLGFFE-W-----TSRLRSHINTGTGVLL 415
DB 362 AORIETTAKAVG-----LPKSRGYTMRDLWKHSDTNTTGTGRIATVPAHGTVLV 409
QY 416 QL 417
DB 410 RV 411
RESULT 13
T39118
probable alpha-galactosidase - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T39118
R:Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, November 1999
A:Reference number: Z21829
A:Accession: T39118
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-436 <HUN>
A:Cross-references: UNIPROT:Q9URZ0; UNIPARC:UPI000000BE8C; EMBL:AL132779; PIDN: CAB60017
A:Experimental source: strain 972h-; cosmid c869
C:Genetics:
A:Gene: SPDB:SPAC869.07c
A:Map position: 1
C:Superfamily: alpha-galactosidase

Query Match 26.6%; Score 609; DB 2; Length 436;
Best Local Similarity 35.8%; Pred. No. 2.3e-44;
Matches 136; Conservative 65; Mismatches 119; Indels 60; Gaps 14;
QY 10 LGCALALRFLALVSWDIPGARALDNGIARTPTMGWLHWRFCMLDQCQEPDSCISEKLF 69
DB 7 LNC-FFLVFLFLPSPDVHGSY---NGLGLKPEQMGMNSWNYACDID-----ESII 52
QY 70 MEMAELMVSSEGWKAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGHGIROLANYVHSGK 128
DB 53 LNNAKAIKEEGLLDLGYEYIVMDDCWSKHERNATTGRLEANPDKFPNGIGSMAXKGLHMG 112
QY 129 LKLGIVADVGNKTCAGPFGSGFYVDIDAQTFADWGVLLKFDGVCYCDLENLA---DGY 184
DB 113 FKGMYSAGKYTCAGPFGSLNHEQIDADTFADWGVVLYKYDNCFNKSGKGVPLISVRY 172
QY 185 KHMSLALNRTGRSIVYS-CEW-PLYMWPFOKPNYTEIROCYCNHWNFNADIDDSWK----- 237
DB 173 KEMSDALNKTGRPIFYSLCQWGEDFVNWG-----NTIANSWRISGDIFFTFSRKDVR 225
QY 238 -----SIKSILDWTSFNOERIVDVAG- PGWNDPDMVLVIGNFGLSNQO 280
DB 226 CPCETIECFALQGDHCSVMNIISKASFSSK---AGMNSGWNDLDSLEVNGSGMSFEY 281
QY 281 VTQMAIWAIPAFLFWSNDRHISPOAKALLOKQVIAINQDPLGKQG---YOLRQCDN 336
DB 282 KTHFTWAILKSPILIGNDVSSPMDKLIVSNKELISINQD-IGTNPAAIWKKKYGYDE 340
QY 337 F-EVWERPLSGLAWAVAMIN 355
DB 341 YIELFSGELSNNDWVAVLN 360

RESULT 14

```

C:Superfamily: alpha-galactosidase
C:keywords: glycoprotein; glycosidase; hydrolase
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-469/Product: alpha-galactosidase MEL #status predicted <MAT>

Query Match      24.3%; Score 557; DB 2; Length 469;
Best Local Similarity 35.3%; Pred. No. 7.4e-40;
Matches 134; Conservative 52; Mismatches 128; Indels 66; Gaps 12;

18 FLALV--SWDIPGARALONGLARTPTMGWLHWRFMCNLDQREPDSCISEKLFMBMAEL 75
5 PFALFFSSTDVLAASPSYNGLGITQMGWDNNSFGCS-----VKELLGTAEK 54
76 MYSEGKQAGYBYLCIDDCWMAQRDSEGRLOADPORFPHGIRQLANYVHSGKLGIIYA 135
55 IVKLGKDLGYNYIILDDCW--SSGRSSNGSLADDSKFPFGMKYVAEQHLNSQLKFGMYS 113
136 DVGNKTCAGFGSGFYVIDIAQTEADGWVDLLKFDGCV---CDSLENLADGYKHWLSAL 191
114 SAGEYTCAGTAGSLGYEDMDAATFASWDVYLKIDNCYNKGEGFTPEISYKRYKAMSDAL 173
192 NRTGRSIVYS-CEWP---LYMPPFOKPNYTEIRQYCNHWRNFADI----- 232
174 NKTGRPIFYSLCNWGQDLTFYWG-----SAISNSWRMSGDVYPQFDRPDRPCPCSG 224
233 ---DDSWK-----SIKSILDWTSFNOERIVDVAGPGGWNDPDLVIQNGFLSNQOVTQMA 285
225 DEYDCSYPGFHCISIMNILKAAPMGQN-----AAPGWNLDLMLEVGVGNNMSDSEVAHFS 280
286 LWAIWAAPLFMGNDLRHISPOAKALLQDKDVIAINQDPLGKQG-----YQLRGDNF--- 337
281 MWAIVKSPLIIGADIDDLKOSLSVYSNPVAVIAINQDVLGTPATRIWKYHVSKDQYGE 340
338 --EVWERPLSGLAWAVAMIN 355
341 EQLWSGFLDNGDHVVALLN 360

Search completed: December 31, 2005, 23:47:42
Job time : 36.5517 secs

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 23:47:17 ; Search time 122.703 Seconds
(without alignments)
1419.967 Million cell updates/sec

Title: US-10-602-220-16

Perfect score: 2293
Sequence: 1 MQLRNPELHGCALALRFLA.....EWSRLRSHINPTGTVLQL 417

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.Main.*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2293	100.0	417	3	US-09-993-059-16
2	2293	100.0	417	4	US-10-103-327-16
3	2293	100.0	417	4	US-10-602-219-16
4	2293	100.0	417	4	US-10-602-220-16
5	2293	100.0	417	5	US-10-851-388-16
6	2293	100.0	417	5	US-10-984-389-16
7	2293	100.0	421	3	US-09-993-059-12
8	2293	100.0	421	4	US-10-103-327-12
9	2293	100.0	421	4	US-10-602-219-12
10	2293	100.0	421	4	US-10-602-220-12
11	2293	100.0	421	5	US-10-851-388-12
12	2293	100.0	421	5	US-10-984-389-12
13	2293	100.0	423	3	US-09-993-059-18
14	2293	100.0	423	4	US-10-103-327-18
15	2293	100.0	423	4	US-10-602-219-18
16	2293	100.0	423	4	US-10-602-220-18
17	2293	100.0	423	5	US-10-851-388-18
18	2293	100.0	423	5	US-10-984-389-18
19	2293	100.0	427	3	US-09-993-059-14
20	2293	100.0	427	4	US-10-103-327-14
21	2293	100.0	427	4	US-10-602-219-14
22	2293	100.0	427	4	US-10-602-220-14
23	2293	100.0	427	5	US-10-851-388-14
24	2293	100.0	427	5	US-10-984-389-14
25	2293	100.0	429	4	US-10-411-037-68
26	2293	100.0	429	4	US-10-411-026-68
27	2293	100.0	429	4	US-10-410-962-68

28	2293	100.0	429	4	US-10-411-049-68	Sequence 68, Appl
29	2293	100.0	429	4	US-10-410-930-68	Sequence 68, Appl
30	2293	100.0	429	4	US-10-410-997-68	Sequence 68, Appl
31	2293	100.0	429	4	US-10-411-012-68	Sequence 68, Appl
32	2293	100.0	429	4	US-10-410-913-68	Sequence 68, Appl
33	2293	100.0	429	5	US-10-370-715B-278	Sequence 278, Appl
34	2293	100.0	429	5	US-10-410-980-68	Sequence 68, Appl
35	2293	100.0	429	5	US-10-410-897-68	Sequence 68, Appl
36	2293	100.0	431	3	US-09-993-059-10	Sequence 10, Appl
37	2293	100.0	431	4	US-10-103-327-10	Sequence 10, Appl
38	2293	100.0	431	4	US-10-602-219-10	Sequence 10, Appl
39	2293	100.0	431	4	US-10-602-220-10	Sequence 10, Appl
40	2293	100.0	431	5	US-10-851-388-10	Sequence 10, Appl
41	2293	100.0	431	5	US-10-984-389-10	Sequence 10, Appl
42	2293	100.0	435	3	US-09-993-059-6	Sequence 6, Appl
43	2293	100.0	435	4	US-10-103-327-6	Sequence 6, Appl
44	2293	100.0	435	4	US-10-602-219-6	Sequence 6, Appl
45	2293	100.0	435	4	US-10-602-220-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-09-993-059-16
; Sequence 16, Application US/09993059
; Publication No. US2002008024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Stephen H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CUPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-16

Query Match	100.0%;	Score	2293;	DB	3;	Length	417;
Best Local Similarity	100.0%;	Pred. No.	9e-220;				
Matches	417;	Conservative	0;	Mismatches	0;	Indels	0;
Qy	1	MQLRNPELHGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWFPMCNLDCEEP	60				
Db	1	MQLRNPELHGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWFPMCNLDCEEP	60				
Qy	61	DSICSEKLFMEALMVSEGMQDAGVEYLCIDCWMAQDSEGRLOADPQRFPHGIRQL	120				
Db	61	DSICSEKLFMEALMVSEGMQDAGVEYLCIDCWMAQDSEGRLOADPQRFPHGIRQL	120				
Qy	121	ANYVHSGKLGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFGDGYCDSLENL	180				
Db	121	ANYVHSGKLGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFGDGYCDSLENL	180				
Qy	181	ADGKMSLALNTGSRISVYSCWPLYMWPQKPNYTEIRQYCNHWNFADIDDSWSIK	240				
Db	181	ADGKMSLALNTGSRISVYSCWPLYMWPQKPNYTEIRQYCNHWNFADIDDSWSIK	240				
Qy	241	SILDWTSFNOCRIVDVAGPGGNDPDMVLVGNFGLSNQVQVOTQALWAAAPLFMSNDL	300				
Db	241	SILDWTSFNOCRIVDVAGPGGNDPDMVLVGNFGLSNQVQVOTQALWAAAPLFMSNDL	300				
Qy	301	RHISPOKALLQDKVIAINQDPLGKGYQLRGQDNFVWVERPLSGLAWAMINRQEIG	360				
Db	301	RHISPOKALLQDKVIAINQDPLGKGYQLRGQDNFVWVERPLSGLAWAMINRQEIG	360				
Qy	361	GPRSYTIAVSLGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL	417				

Db 361 GPRSYYTIAVASLKGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 2

US-10-103-327-16
; Sequence 16, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-16

Query Match 100.0%; Score 2293; DB 4; Length 417;
Best Local Similarity 100.0%; Pred. No. 9e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MQLRNPGLHGCALALRFLALVSNDIPGARALDGLARTPTMGWLHWRFCMNLDCOEEP	60
Db	1	MQLRNPGLHGCALALRFLALVSNDIPGARALDGLARTPTMGWLHWRFCMNLDCOEEP	60
Qy	61	DSICSEKLFMEAEMLVSEGWKADAGYEYLCIDDCWAPQDSEGRLOADPQRPFGIRQL	120
Db	61	DSICSEKLFMEAEMLVSEGWKADAGYEYLCIDDCWAPQDSEGRLOADPQRPFGIRQL	120
Qy	121	ANYVHSGKLGKIYADVGNKTCAGFPGSFGYYIDIAQTADGWVDLLKFDGCGYCDLSLENL	180
Db	121	ANYVHSGKLGKIYADVGNKTCAGFPGSFGYYIDIAQTADGWVDLLKFDGCGYCDLSLENL	180
Qy	181	ADGYKMSLALNRTGRSIVYSCWPLYMWPFPKPNYTEIRQYCNHWRNFADIDDSWKSIIK	240
Db	181	ADGYKMSLALNRTGRSIVYSCWPLYMWPFPKPNYTEIRQYCNHWRNFADIDDSWKSIIK	240
Qy	241	SILDWTSFNQERIVDVAGPGGNDPDMVLVGNFGLSNQOVTOMALWAIWAAPLFWMSNDL	300
Db	241	SILDWTSFNQERIVDVAGPGGNDPDMVLVGNFGLSNQOVTOMALWAIWAAPLFWMSNDL	300
Qy	301	RHISPOKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQEIG	360
Db	301	RHISPOKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQEIG	360
Qy	361	GPRSYYTIAVASLKGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL	417
Db	361	GPRSYYTIAVASLKGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL	417

RESULT 3

US-10-602-219-16
; Sequence 16, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: POGUE, Gregory P.
; APPLICANT: ERWIN, Robert L.
; APPLICANT: GRILL, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219

; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-602-219-16

Query Match	100.0%;	Score 2293;	DB 4;	Length 417;
Best Local Similarity	100.0%;	Pred. No. 9e-220;		
Matches	417;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Qy	1	MQLRNPGLHGCALALRFLALVSNDIPGARALDGLARTPTMGWLHWRFCMNLDCOEEP	60	
Db	1	MQLRNPGLHGCALALRFLALVSNDIPGARALDGLARTPTMGWLHWRFCMNLDCOEEP	60	
Qy	61	DSICSEKLFMEAEMLVSEGWKADAGYEYLCIDDCWAPQDSEGRLOADPQRPFGIRQL	120	
Db	61	DSICSEKLFMEAEMLVSEGWKADAGYEYLCIDDCWAPQDSEGRLOADPQRPFGIRQL	120	
Qy	121	ANYVHSGKLGKIYADVGNKTCAGFPGSFGYYIDIAQTADGWVDLLKFDGCGYCDLSLENL	180	
Db	121	ANYVHSGKLGKIYADVGNKTCAGFPGSFGYYIDIAQTADGWVDLLKFDGCGYCDLSLENL	180	
Qy	181	ADGYKMSLALNRTGRSIVYSCWPLYMWPFPKPNYTEIRQYCNHWRNFADIDDSWKSIIK	240	
Db	181	ADGYKMSLALNRTGRSIVYSCWPLYMWPFPKPNYTEIRQYCNHWRNFADIDDSWKSIIK	240	
Qy	241	SILDWTSFNQERIVDVAGPGGNDPDMVLVGNFGLSNQOVTOMALWAIWAAPLFWMSNDL	300	
Db	241	SILDWTSFNQERIVDVAGPGGNDPDMVLVGNFGLSNQOVTOMALWAIWAAPLFWMSNDL	300	
Qy	301	RHISPOKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQEIG	360	
Db	301	RHISPOKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQEIG	360	
Qy	361	GPRSYYTIAVASLKGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL	417	
Db	361	GPRSYYTIAVASLKGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL	417	

RESULT 4

US-10-602-220-16
; Sequence 16, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; APPLICANT: POGUE, Gregory P.
; APPLICANT: ERWIN, Robert L.

APPLICANT: Grill, Laurence K.
TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
FILE REFERENCE: LSEC-0087-CF07B
CURRENT APPLICATION NUMBER: US/10/602,220
CURRENT FILING DATE: 2003-06-23
PRIORITY APPLICATION NUMBER: 09/993,059
PRIORITY FILING DATE: 2001-11-13
PRIORITY APPLICATION NUMBER: 09/626,127
PRIORITY FILING DATE: 2000-07-26
PRIORITY APPLICATION NUMBER: 09/316,572
PRIORITY FILING DATE: 1999-05-21
PRIORITY APPLICATION NUMBER: 08/324,003
PRIORITY FILING DATE: 1994-10-14
PRIORITY APPLICATION NUMBER: 08/176,414
PRIORITY FILING DATE: 1993-12-29
PRIORITY APPLICATION NUMBER: 07/997,733
PRIORITY FILING DATE: 1992-12-30
PRIORITY APPLICATION NUMBER: 08/184,237
PRIORITY FILING DATE: 1994-01-19
PRIORITY APPLICATION NUMBER: 07/923,692
PRIORITY FILING DATE: 1992-07-31
PRIORITY APPLICATION NUMBER: 07/600,244
PRIORITY FILING DATE: 1990-10-23
PRIORITY APPLICATION NUMBER: 07/641,617
PRIORITY FILING DATE: 1991-01-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn version 3.2
SEQ ID NO 16
LENGTH: 417
TYPE: PRT
ORGANISM: Homo sapiens
US-10-602-220-16

Query Match 100.0%; Score 2293; DB 4; Length 417;
Best Local Similarity 100.0%; Pred. No. 9e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPGLHGCALALRFLALVSDIPGARALDNLARTPTMGWLHWFRCNLDCCQEEP 60
Db 1 MQLRNPGLHGCALALRFLALVSDIPGARALDNLARTPTMGWLHWFRCNLDCCQEEP 60

Qy 61 DSCISEKLFMEAEMLVSEGWKAGYEYLCIDDCWMAPOQDSEGRLOADPQPFPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKAGYEYLCIDDCWMAPOQDSEGRLOADPQPFPHGIRQL 120

Qy 121 ANYVHSGKLGIYADVGNKTCAGFGSGYDYIDAQTFADMGVDLLKFGDGCYDLSLENL 180
Db 121 ANYVHSGKLGIYADVGNKTCAGFGSGYDYIDAQTFADMGVDLLKFGDGCYDLSLENL 180

Qy 181 ADGKXMSLALNRTGRSIVVSCWPLVYMFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKXMSLALNRTGRSIVVSCWPLVYMFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240

Qy 241 SILDWTSENGERIVDVAGPGWNPDMVLVINGFGLSNQVQVQWALWAIMAAPLFMSNDL 300
Db 241 SILDWTSENGERIVDVAGPGWNPDMVLVINGFGLSNQVQVQWALWAIMAAPLFMSNDL 300

Qy 301 RHISPOKALLQDKVDVAINQDPLGKQGYQLROGDNFEVWERPLSGLAWAVAMINROEIG 360
Db 301 RHISPOKALLQDKVDVAINQDPLGKQGYQLROGDNFEVWERPLSGLAWAVAMINROEIG 360

Qy 361 GPRSYTIAVSLGKGVACNPACFTITQLLPVKRLGFEYWTSLRSHINPTGTLLQL 417
Db 361 GPRSYTIAVSLGKGVACNPACFTITQLLPVKRLGFEYWTSLRSHINPTGTLLQL 417

RESULT 5
US-10-851-388-16
Sequence 16, Application US/10851388
Publication No. US20040234516A1
GENERAL INFORMATION:
APPLICANT: GARGER, Stephen A.

APPLICANT: TURPEN, Thomas H.
APPLICANT: KUMAGAI, Monto H.
TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
FILE REFERENCE: 008010087CPUS06
CURRENT APPLICATION NUMBER: US/10/851,388
CURRENT FILING DATE: 2004-05-21
PRIORITY APPLICATION NUMBER: US/09/993,059
PRIORITY FILING DATE: 2001-11-13
NUMBER OF SEQ ID NOS: 37
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16
LENGTH: 417
TYPE: PRT
ORGANISM: Homo sapiens
US-10-851-388-16

Query Match 100.0%; Score 2293; DB 5; Length 417;
Best Local Similarity 100.0%; Pred. No. 9e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPGLHGCALALRFLALVSDIPGARALDNLARTPTMGWLHWFRCNLDCCQEEP 60
Db 1 MQLRNPGLHGCALALRFLALVSDIPGARALDNLARTPTMGWLHWFRCNLDCCQEEP 60

Qy 61 DSCISEKLFMEAEMLVSEGWKAGYEYLCIDDCWMAPOQDSEGRLOADPQPFPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKAGYEYLCIDDCWMAPOQDSEGRLOADPQPFPHGIRQL 120

Qy 121 ANYVHSGKLGIYADVGNKTCAGFGSGYDYIDAQTFADMGVDLLKFGDGCYDLSLENL 180
Db 121 ANYVHSGKLGIYADVGNKTCAGFGSGYDYIDAQTFADMGVDLLKFGDGCYDLSLENL 180

Qy 181 ADGKXMSLALNRTGRSIVVSCWPLVYMFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKXMSLALNRTGRSIVVSCWPLVYMFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240

Qy 241 SILDWTSENGERIVDVAGPGWNPDMVLVINGFGLSNQVQVQWALWAIMAAPLFMSNDL 300
Db 241 SILDWTSENGERIVDVAGPGWNPDMVLVINGFGLSNQVQVQWALWAIMAAPLFMSNDL 300

Qy 301 RHISPOKALLQDKVDVAINQDPLGKQGYQLROGDNFEVWERPLSGLAWAVAMINROEIG 360
Db 301 RHISPOKALLQDKVDVAINQDPLGKQGYQLROGDNFEVWERPLSGLAWAVAMINROEIG 360

Qy 361 GPRSYTIAVSLGKGVACNPACFTITQLLPVKRLGFEYWTSLRSHINPTGTLLQL 417
Db 361 GPRSYTIAVSLGKGVACNPACFTITQLLPVKRLGFEYWTSLRSHINPTGTLLQL 417

RESULT 6
US-10-984-389-16
Sequence 16, Application US/10984389
Publication No. US20050125859A1
GENERAL INFORMATION:
APPLICANT: GARGER, Stephen A.
APPLICANT: TURPEN, Thomas H.
APPLICANT: KUMAGAI, Monto H.
TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
FILE REFERENCE: 008010087CPUS06
CURRENT APPLICATION NUMBER: US/10/984,389
CURRENT FILING DATE: 2004-11-08
PRIORITY APPLICATION NUMBER: US/09/993,059
PRIORITY FILING DATE: 2001-11-13
NUMBER OF SEQ ID NOS: 37
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16
LENGTH: 417
TYPE: PRT
ORGANISM: Homo sapiens
US-10-984-389-16

```
Query Match      100.0%; Score 2293; DB 5; Length 417;
Best Local Similarity 100.0%; Pred. No. 9e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPGLHGCALALRFLALVSDIPGARALDNGLARTPTMGLHWRFCMNLDCQEEP 60
Db 1 MQLRNPGLHGCALALRFLALVSDIPGARALDNGLARTPTMGLHWRFCMNLDCQEEP 60

Qy 61 DSCISEKLFMEAMALVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAMALVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120

Qy 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180

Qy 181 ADGKHMSLALNRTGRSIVVSCWPLYMPPQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKHMSLALNRTGRSIVVSCWPLYMPPQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240

Qy 241 SILDWTSEFNOERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQMALWAIWAAPLFMSNDL 300
Db 241 SILDWTSEFNOERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQMALWAIWAAPLFMSNDL 300

Qy 301 RHISPOKALLQDKDVIATINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360
Db 301 RHISPOKALLQDKDVIATINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360

Qy 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 417
Db 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 417

RESULT 7
US-09-993-059-12
; Sequence 12, Application US/09993059
; Publication No. US2002008024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-12

Query Match      100.0%; Score 2293; DB 3; Length 421;
Best Local Similarity 100.0%; Pred. No. 9.1e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPGLHGCALALRFLALVSDIPGARALDNGLARTPTMGLHWRFCMNLDCQEEP 60
Db 1 MQLRNPGLHGCALALRFLALVSDIPGARALDNGLARTPTMGLHWRFCMNLDCQEEP 60

Qy 61 DSCISEKLFMEAMALVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAMALVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120

Qy 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180

Qy 181 ADGKHMSLALNRTGRSIVVSCWPLYMPPQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKHMSLALNRTGRSIVVSCWPLYMPPQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
```

```
Qy 241 SILDWTSEFNOERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQMALWAIWAAPLFMSNDL 300
Db 241 SILDWTSEFNOERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQMALWAIWAAPLFMSNDL 300

Qy 301 RHISPOKALLQDKDVIATINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360
Db 301 RHISPOKALLQDKDVIATINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360

Qy 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 417
Db 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 417

RESULT 8
US-10-103-327-12
; Sequence 12, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/10/103,327
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-12

Query Match      100.0%; Score 2293; DB 4; Length 421;
Best Local Similarity 100.0%; Pred. No. 9.1e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPGLHGCALALRFLALVSDIPGARALDNGLARTPTMGLHWRFCMNLDCQEEP 60
Db 1 MQLRNPGLHGCALALRFLALVSDIPGARALDNGLARTPTMGLHWRFCMNLDCQEEP 60

Qy 61 DSCISEKLFMEAMALVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAMALVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120

Qy 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180

Qy 181 ADGKHMSLALNRTGRSIVVSCWPLYMPPQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKHMSLALNRTGRSIVVSCWPLYMPPQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240

Qy 241 SILDWTSEFNOERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQMALWAIWAAPLFMSNDL 300
Db 241 SILDWTSEFNOERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQMALWAIWAAPLFMSNDL 300

Qy 301 RHISPOKALLQDKDVIATINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360
Db 301 RHISPOKALLQDKDVIATINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360

Qy 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 417
Db 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 417

RESULT 9
US-10-602-219-12
; Sequence 12, Application US/10602219
```

Publication No. US20040016021A1
GENERAL INFORMATION:
APPLICANT: Large Scale Biology Corporation
APPLICANT: Turpen, Thomas H.
APPLICANT: Pogue, Gregory P.
APPLICANT: Erwin, Robert L.
APPLICANT: Grill, Laurence K.
TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
FILE REFERENCE: LSBC-0087-CF09B
CURRENT APPLICATION NUMBER: US/10/602,219
CURRENT FILING DATE: 2003-06-23
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: 09/993,059
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 09/316,572
PRIOR FILING DATE: 1999-05-21
PRIOR APPLICATION NUMBER: 08/324,003
PRIOR FILING DATE: 1994-10-14
PRIOR APPLICATION NUMBER: 08/176,414
PRIOR FILING DATE: 1993-12-29
PRIOR APPLICATION NUMBER: 07/997,733
PRIOR FILING DATE: 1992-12-30
PRIOR APPLICATION NUMBER: 07/923,692
PRIOR FILING DATE: 1990-10-22
PRIOR APPLICATION NUMBER: 07/641,617
PRIOR FILING DATE: 1991-01-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn version 3.2
SEQ ID NO 12
LENGTH: 421
TYPE: PRT
ORGANISM: Homo sapiens
US-10-602-219-12

Query Match 100.0%; Score 2293; DB 4; Length 421;
Best Local Similarity 100.0%; Pred. No. 9.1e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MOLNPELHGCALALRFLALVSDIPCARALDNLGLARTPTMGWLHWRFCNLDCCQEEP 60
Db 1 MOLNPELHGCALALRFLALVSDIPCARALDNLGLARTPTMGWLHWRFCNLDCCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAPOQDSEGRLOADPQPFPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAPOQDSEGRLOADPQPFPHGIRQL 120
Qy 121 ANYVHSGKLGIYADVGNKTCAGFPGSFGYDIDAQTADGWVDLLKFGCYCDSLENL 180
Db 121 ANYVHSGKLGIYADVGNKTCAGFPGSFGYDIDAQTADGWVDLLKFGCYCDSLENL 180
Qy 181 ADGKMSLALNRGSRIVSCWPLVWPFQKPNYTEIRQYCNHWNFPADIDDSWSIK 240
Db 181 ADGKMSLALNRGSRIVSCWPLVWPFQKPNYTEIRQYCNHWNFPADIDDSWSIK 240
Qy 241 SILDWTSFQERIVDVAGPGGNDPDLVIGNFGLSNQVQTQALWAIWAAPLFMSNDL 300
Db 241 SILDWTSFQERIVDVAGPGGNDPDLVIGNFGLSNQVQTQALWAIWAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360
Qy 361 GPRSYTIAVSLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTULLQL 417
Db 361 GPRSYTIAVSLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTULLQL 417

RESULT 10
US-10-602-220-12
Sequence 12, Application US/10602220
Publication No. US20040023281A1
GENERAL INFORMATION:
APPLICANT: Large Scale Biology Corporation
APPLICANT: Turpen, Thomas H.
APPLICANT: Kumagai, Monto H.
APPLICANT: Pogue, Gregory P.
APPLICANT: Erwin, Robert L.
APPLICANT: Grill, Laurence K.
TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
FILE REFERENCE: LSBC-0087-CF07B
CURRENT APPLICATION NUMBER: US/10/602,220
CURRENT FILING DATE: 2003-06-23
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: 09/993,059
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 09/316,572
PRIOR FILING DATE: 1999-05-21
PRIOR APPLICATION NUMBER: 08/324,003
PRIOR FILING DATE: 1994-10-14
PRIOR APPLICATION NUMBER: 08/176,414
PRIOR FILING DATE: 1993-12-29
PRIOR APPLICATION NUMBER: 07/997,733
PRIOR FILING DATE: 1992-12-30
PRIOR APPLICATION NUMBER: 08/184,237
PRIOR FILING DATE: 1994-01-19
PRIOR APPLICATION NUMBER: 07/923,692
PRIOR FILING DATE: 1992-07-31
PRIOR APPLICATION NUMBER: 07/600,244
PRIOR FILING DATE: 1990-10-22
PRIOR APPLICATION NUMBER: 07/641,617
PRIOR FILING DATE: 1991-01-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn version 3.2
SEQ ID NO 12
LENGTH: 421
TYPE: PRT
ORGANISM: Homo sapiens
US-10-602-220-12

Query Match 100.0%; Score 2293; DB 4; Length 421;
Best Local Similarity 100.0%; Pred. No. 9.1e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MOLNPELHGCALALRFLALVSDIPCARALDNLGLARTPTMGWLHWRFCNLDCCQEEP 60
Db 1 MOLNPELHGCALALRFLALVSDIPCARALDNLGLARTPTMGWLHWRFCNLDCCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAPOQDSEGRLOADPQPFPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAPOQDSEGRLOADPQPFPHGIRQL 120
Qy 121 ANYVHSGKLGIYADVGNKTCAGFPGSFGYDIDAQTADGWVDLLKFGCYCDSLENL 180
Db 121 ANYVHSGKLGIYADVGNKTCAGFPGSFGYDIDAQTADGWVDLLKFGCYCDSLENL 180
Qy 181 ADGKMSLALNRGSRIVSCWPLVWPFQKPNYTEIRQYCNHWNFPADIDDSWSIK 240
Db 181 ADGKMSLALNRGSRIVSCWPLVWPFQKPNYTEIRQYCNHWNFPADIDDSWSIK 240
Qy 241 SILDWTSFQERIVDVAGPGGNDPDLVIGNFGLSNQVQTQALWAIWAAPLFMSNDL 300
Db 241 SILDWTSFQERIVDVAGPGGNDPDLVIGNFGLSNQVQTQALWAIWAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQGDNFVWERPLSGLAWAVAMINRQIEG 360
Qy 361 GPRSYTIAVSLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTULLQL 417

```
Db 361 GPRSYTIAVSLGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
|||||
RESULT 11
US-10-851-388-12
; Sequence 12, Application US/10851388
; Publication No. US20040234516A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851,388
; PRIOR FILING DATE: 2004-05-21
; PRIOR FILING DATE: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-851-388-12

Query Match 100.0%; Score 2293; DB 5; Length 421;
Best Local Similarity 100.0%; Pred. No. 9.1e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPELHGCALALRFLALVSWDIPGARALDNLGLARTPTMGLHWRFCNLDCCQEEP 60
Db 1 MQLRNPELHGCALALRFLALVSWDIPGARALDNLGLARTPTMGLHWRFCNLDCCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGFYDIDAQTFADWGVDDLKFGDCCDSLENL 180
Db 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGFYDIDAQTFADWGVDDLKFGDCCDSLENL 180
Qy 181 ADGKHSALNRTGRSIVYSCWPLVMFPQKPNYTEIRQYCNHWRNFADIDDSWSIK 240
Db 181 ADGKHSALNRTGRSIVYSCWPLVMFPQKPNYTEIRQYCNHWRNFADIDDSWSIK 240
Qy 241 SILDWTSFNQERIVDVAGPGGWNDDPMLVIGNFGLSNQQTQMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSFNQERIVDVAGPGGWNDDPMLVIGNFGLSNQQTQMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQDGFVWERPLSGLAWAVAMINRQEI 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQDGFVWERPLSGLAWAVAMINRQEI 360
Qy 361 GPRSYTIAVSLGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
Db 361 GPRSYTIAVSLGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 12
US-10-851-388-12
; Sequence 12, Application US/10984389
; Publication No. US20050125859A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/984,389
; CURRENT FILING DATE: 2004-11-08
; SEQ ID NO 18
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-984-389-12

Query Match 100.0%; Score 2293; DB 3; Length 423;
Best Local Similarity 100.0%; Pred. No. 9.2e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPELHGCALALRFLALVSWDIPGARALDNLGLARTPTMGLHWRFCNLDCCQEEP 60
Db 1 MQLRNPELHGCALALRFLALVSWDIPGARALDNLGLARTPTMGLHWRFCNLDCCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGFYDIDAQTFADWGVDDLKFGDCCDSLENL 180
Db 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGFYDIDAQTFADWGVDDLKFGDCCDSLENL 180
Qy 181 ADGKHSALNRTGRSIVYSCWPLVMFPQKPNYTEIRQYCNHWRNFADIDDSWSIK 240
Db 181 ADGKHSALNRTGRSIVYSCWPLVMFPQKPNYTEIRQYCNHWRNFADIDDSWSIK 240
Qy 241 SILDWTSFNQERIVDVAGPGGWNDDPMLVIGNFGLSNQQTQMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSFNQERIVDVAGPGGWNDDPMLVIGNFGLSNQQTQMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQDGFVWERPLSGLAWAVAMINRQEI 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQDGFVWERPLSGLAWAVAMINRQEI 360
Qy 361 GPRSYTIAVSLGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
Db 361 GPRSYTIAVSLGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
```

```
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-984-389-12

Query Match 100.0%; Score 2293; DB 5; Length 421;
Best Local Similarity 100.0%; Pred. No. 9.1e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPELHGCALALRFLALVSWDIPGARALDNLGLARTPTMGLHWRFCNLDCCQEEP 60
Db 1 MQLRNPELHGCALALRFLALVSWDIPGARALDNLGLARTPTMGLHWRFCNLDCCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGFYDIDAQTFADWGVDDLKFGDCCDSLENL 180
Db 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGFYDIDAQTFADWGVDDLKFGDCCDSLENL 180
Qy 181 ADGKHSALNRTGRSIVYSCWPLVMFPQKPNYTEIRQYCNHWRNFADIDDSWSIK 240
Db 181 ADGKHSALNRTGRSIVYSCWPLVMFPQKPNYTEIRQYCNHWRNFADIDDSWSIK 240
Qy 241 SILDWTSFNQERIVDVAGPGGWNDDPMLVIGNFGLSNQQTQMALWAIMAAPLFMSNDL 300
Db 241 SILDWTSFNQERIVDVAGPGGWNDDPMLVIGNFGLSNQQTQMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQDGFVWERPLSGLAWAVAMINRQEI 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQDGFVWERPLSGLAWAVAMINRQEI 360
Qy 361 GPRSYTIAVSLGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
Db 361 GPRSYTIAVSLGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 13
US-09-993-059-18
; Sequence 18, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-18

Query Match 100.0%; Score 2293; DB 3; Length 423;
Best Local Similarity 100.0%; Pred. No. 9.2e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPELHGCALALRFLALVSWDIPGARALDNLGLARTPTMGLHWRFCNLDCCQEEP 60
Db 1 MQLRNPELHGCALALRFLALVSWDIPGARALDNLGLARTPTMGLHWRFCNLDCCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
```

```
Db 61 DSCISEKLFMEALMVSEGWKDAYEYLCIDDCWMAQRDSEGRQLQADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGIGIYADVGNKTCAGFPGSFGYYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGIGIYADVGNKTCAGFPGSFGYYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
Qy 181 ADGKXMSLALNRTGRSIVTSCWPLVYMWFPQKNTYIROYCNHWNFNADIDDSWSKSIK 240
Db 181 ADGKXMSLALNRTGRSIVTSCWPLVYMWFPQKNTYIROYCNHWNFNADIDDSWSKSIK 240
Qy 241 SILDWTSFNOERIVDVAGPGGNDPDMVLVGNFGLSNQVQVTOALWAIMAAPLFMSNDL 300
Db 241 SILDWTSFNOERIVDVAGPGGNDPDMVLVGNFGLSNQVQVTOALWAIMAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKDVTAINDPGLKQGYQLRQDGFVWVERPLSLGLAWAVAMINROEIG 360
Db 301 RHISPOKALLQDKDVTAINDPGLKQGYQLRQDGFVWVERPLSLGLAWAVAMINROEIG 360
Qy 361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
Db 361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 14
US-10-103-327-18
; Sequence 18, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-18

Query Match 100.0%; Score 2293; DB 4; Length 423;
Best Local Similarity 100.0%; Pred. No. 9.2e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Db 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Qy 61 DSCISEKLFMEALMVSEGWKDAYEYLCIDDCWMAQRDSEGRQLQADPQRPFGIRQL 120
Db 61 DSCISEKLFMEALMVSEGWKDAYEYLCIDDCWMAQRDSEGRQLQADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGIGIYADVGNKTCAGFPGSFGYYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGIGIYADVGNKTCAGFPGSFGYYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
Qy 181 ADGKXMSLALNRTGRSIVTSCWPLVYMWFPQKNTYIROYCNHWNFNADIDDSWSKSIK 240
Db 181 ADGKXMSLALNRTGRSIVTSCWPLVYMWFPQKNTYIROYCNHWNFNADIDDSWSKSIK 240
Qy 241 SILDWTSFNOERIVDVAGPGGNDPDMVLVGNFGLSNQVQVTOALWAIMAAPLFMSNDL 300
Db 241 SILDWTSFNOERIVDVAGPGGNDPDMVLVGNFGLSNQVQVTOALWAIMAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKDVTAINDPGLKQGYQLRQDGFVWVERPLSLGLAWAVAMINROEIG 360
```

```
Db 301 RHISPOKALLQDKDVTAINDPGLKQGYQLRQDGFVWVERPLSLGLAWAVAMINROEIG 360
Qy 361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
Db 361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 15
US-10-602-219-18
; Sequence 18, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 18
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-602-219-18

Query Match 100.0%; Score 2293; DB 4; Length 423;
Best Local Similarity 100.0%; Pred. No. 9.2e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Db 1 MOLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Qy 61 DSCISEKLFMEALMVSEGWKDAYEYLCIDDCWMAQRDSEGRQLQADPQRPFGIRQL 120
Db 61 DSCISEKLFMEALMVSEGWKDAYEYLCIDDCWMAQRDSEGRQLQADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGIGIYADVGNKTCAGFPGSFGYYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGIGIYADVGNKTCAGFPGSFGYYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
Qy 181 ADGKXMSLALNRTGRSIVTSCWPLVYMWFPQKNTYIROYCNHWNFNADIDDSWSKSIK 240
Db 181 ADGKXMSLALNRTGRSIVTSCWPLVYMWFPQKNTYIROYCNHWNFNADIDDSWSKSIK 240
Qy 241 SILDWTSFNOERIVDVAGPGGNDPDMVLVGNFGLSNQVQVTOALWAIMAAPLFMSNDL 300
Db 241 SILDWTSFNOERIVDVAGPGGNDPDMVLVGNFGLSNQVQVTOALWAIMAAPLFMSNDL 300
```

Qy	301	RHISPOAKALLQDKDVIAINQDPLGKQGYQLROGDNFEVWERPLSGLAWAVAMINRQEIG	360
Db	301	RHISPOAKALLQDKDVIAINQDPLGKQGYQLROGDNFEVWERPLSGLAWAVAMINRQEIG	360
Qy	361	GPRSYTTIIVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL	417
Db	361	GPRSYTTIIVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL	417

Search completed: January 1, 2006, 00:29:24
Job time : 124.703 secs

Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.

This Page Blank (uspto)